LARGE DEVIATION PRINCIPLES FOR EMPIRICAL MEASURES OF COLOURED RANDOM GRAPHS

By Kwabena Doku-Amponsah and Peter Mörters

University of Bath

Abstract. For any finite coloured graph we define the *empirical neighbourhood measure*, which counts the number of vertices of a given colour connected to a given number of vertices of each colour, and the *empirical pair measure*, which counts the number of edges connecting each pair of colours. For a class of models of sparse coloured random graphs, we prove large deviation principles for these empirical measures in the weak topology. The rate functions governing our large deviation principles can be expressed explicitly in terms of relative entropies. We derive a large deviation principle for the degree distribution of Erdős-Rényi graphs near criticality.

1. Introduction

In this paper we study a random graph model where each vertex of the graph carries a random symbol, spin or colour. The easiest model of this kind is that of an Erdős-Rényi graph where additionally each vertex is equipped with an independently chosen colour. The more general models of coloured random graphs we consider here allow for a dependence between colour and connectivity of the vertices.

With each coloured graph we associate its *empirical neighbourhood measure*, which records the number of vertices of a given colour with a given number of adjacent vertices of each colour. From this quantity one can derive a host of important characteristics of the coloured graph, like its degree distribution, the number of edges linking two given colours, or the number of isolated vertices of any colour. The aim of this paper is to derive a large deviation principle for the empirical neighbourhood measure.

Our motivation for this project is twofold. On the one hand one may consider the coloured random graphs as a very simple random model of *networked data*. The data is described as a text of fixed length, consisting of words chosen from a finite dictionary, together with a random number of unoriented edges or links connecting the words. Large deviation results for the empirical neighbourhood measure permit the calculation of the asymptotic number of bits needed to transmit a large amount of such data with arbitrarily small error probability, see [DA06a, DA06b] where this idea is followed up.

On the other hand we are working towards understanding simple models of *statistical mechanics* defined on random graphs. Here, typically, the colours of the vertices are interpreted as spins, taken from a finite set of possibilities, and the Hamiltonian of the system is an integral of some function with respect to the empirical neighbourhood measure. As a very simple example we provide the annealed asymptotics of the random partition function for the Ising model on an Erdős-Rényi graph, as the graph size goes to infinity.

To be more specific about our model, we consider coloured random graphs constructed as follows: In the first step each of n fixed vertices independently gets a colour, chosen according to some law μ on the finite set $\mathcal{X}$ of colours. In the second step we connect any pair of vertices independently with a probability $p(a, b)$ depending on the colours $a, b \in \mathcal{X}$ of the two vertices. This model, which comprises the simple Erdős-Rényi graph with independent colours as a special case, was introduced by Penman in his thesis [Pe98], see [CP03] for an exposition.

Our main concern in this paper are asymptotic results when the number n of vertices go to infinity, while the connection probabilities go to zero of order $1/n$. This leads to an average number of edges of order n, the *near critical* or *sparse* case. Our methods also allow the study of the sub- and supercritical regimes. Some results on these cases are discussed in [DA06a, DA06b].

Recall that a *rate function* is a non-constant, lower semicontinuous function I from a polish space $\mathcal{M}$ into $[0, \infty]$, it is called *good* if the level sets $\{I(m) \le x\}$ are compact for every $x \in [0, \infty)$. A functional M from the set of finite coloured graphs to $\mathcal{M}$ is said to satisfy a *large deviation principle* with rate function I if, for all Borel sets $B \subset \mathcal{M}$,

$$- \inf_{m \in \mathrm{int}\, B} I(m) \le \liminf_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}_n\big\{M(X) \in B\big\}$$
$$\le \limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}_n\big\{M(X) \in B\big\} \le - \inf_{m \in \mathrm{cl}\, B} I(m),$$

where X under $\mathbb{P}_n$ is a coloured random graph with n vertices and int B and cl B refer to the interior, resp. closure, of the set B.

Apart from the empirical neighbourhood measure defined above, we also consider the *empirical pair measure*, which counts the number of edges connecting any given pair of colours, and the *empirical colour measure*, which simply counts the number of vertices of any given colour. The main result of this paper is a joint large deviation principle for the empirical neighbourhood measure and the empirical pair measure of a coloured random graph in the weak topology, see Theorem 2.1. In the course of the proof of this principle, two further large deviation principles are established: A large deviation principle for the empirical neighbourhood measure conditioned to have a given empirical pair and colour measure, see Theorem 2.5, and a joint large deviation principle for the empirical colour measure and the empirical pair measure, see Theorem 2.3. For all these principles we obtain a completely explicit rate function given in terms of relative entropies.

As an example, we consider the Erdős-Rényi graph model on n vertices, where edges are inserted with probability $p_n \in [0, 1]$ independently for any pair of vertices. We assume that $np_n \to c \in (0, 1)$. From our main result we derive a large deviation principle for the degree distribution, see Corollary 2.2. This example seems to be new in this explicit form.

2. Statement of the results

Let V be a fixed set of n vertices, say $V = \{1, \ldots, n\}$ and denote by $\mathcal{G}_n$ the set of all (simple) graphs with vertex set $V = \{1, \ldots, n\}$ and edge set $E \subset \mathcal{E} := \big\{(u, v) \in V \times V : u < v\big\}$, where the *formal* ordering of edges is introduced as a means to simply describe *unordered* edges. Note that for all n, we have $0 \le |E| \le \tfrac{1}{2} n(n-1)$. Let $\mathcal{X}$ be a finite alphabet or colour set $\mathcal{X}$ and denote by $\mathcal{G}_n(\mathcal{X})$ be the set of all coloured graphs with colour set $\mathcal{X}$ and n vertices.

Given a symmetric function $p_n \colon \mathcal{X} \times \mathcal{X} \to [0, 1]$ and a probability measure μ on $\mathcal{X}$ we may define the *randomly coloured random graph* or simply *coloured random graph* X with n vertices as follows: Assign to each vertex $v \in V$ colour $X(v)$ independently according to the *colour law* μ. Given the colours, we connect any two vertices $u, v \in V$, independently of everything else, with *connection*

probability $p_n(X(u), X(v))$. We always consider $X = ((X(v) : v \in V), E)$ under the joint law of graph and colour and interpret X as coloured random graph.

We are interested in the properties of the randomly coloured graphs for large n in the *sparse* or *near critical* case, i.e. we assume that the connection probabilities satisfy $np_n(a, b) \to C(a, b)$ for all $a, b \in \mathcal{X}$, where $C \colon \mathcal{X} \times \mathcal{X} \to [0, \infty)$ is a symmetric function, which is not identically equal to zero.

To fix some notation, for any finite or countable set $\mathcal{Y}$ we denote by $\mathcal{M}(\mathcal{Y})$ the space of probability measures, and by $\tilde{\mathcal{M}}(\mathcal{Y})$ the space of finite measures on $\mathcal{Y}$, both endowed with the weak topology. By $\mathcal{N}(\mathcal{Y})$ we denote the space of counting measures on $\mathcal{Y}$, i.e. those measures taking values in $\mathbb{N} \cup \{0\}$, endowed with the discrete topology. Finally, we denote by $\tilde{\mathcal{M}}_*(\mathcal{Y} \times \mathcal{Y})$ the subspace of symmetric measures in $\tilde{\mathcal{M}}(\mathcal{Y} \times \mathcal{Y})$.

With any coloured graph $X = ((X(v) : v \in V), E)$ with n vertices we associate a probability measure, the *empirical colour measure* $L^1 \in \mathcal{M}(\mathcal{X})$, by

$$L^1(a) := \frac{1}{n} \sum_{v \in V} \delta_{X(v)}(a), \quad \text{for } a \in \mathcal{X},$$

and a symmetric finite measure, the *empirical pair measure* $L^2 \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$, by

$$L^2(a, b) := \frac{1}{n} \sum_{(u,v) \in E} [\delta_{(X(v), X(u))} + \delta_{(X(u), X(v))}](a, b), \quad \text{for } a, b \in \mathcal{X}.$$

The total mass $\|L^2\|$ of L^2, is $2|E|/n$. Finally we define a further probability measure, the *empirical neighbourhood measure* $M \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$, by

$$M(a, \ell) := \frac{1}{n} \sum_{v \in V} \delta_{(X(v), L(v))}(a, \ell), \quad \text{for } (a, \ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X}),$$

where $L(v) = (l^v(b), b \in \mathcal{X})$ and $l^v(b)$ is the number of vertices of colour b connected to vertex v. For every $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ let ν_1, ν_2 be the $\mathcal{X}$-marginal, respectively the $\mathcal{N}(\mathcal{X})$−marginal, of the measure ν. Moreover, we define a measure $\langle \nu(\cdot, \ell), \ell(\cdot) \rangle \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ by

$$\langle \nu(\cdot, \ell), \ell(\cdot) \rangle(a, b) := \sum_{\ell \in \mathcal{N}(\mathcal{X})} \nu(a, \ell)\ell(b), \quad \text{for } a, b \in \mathcal{X}.$$

Define the function $\Phi \colon \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) \to \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ by $\Phi(\nu) = (\nu_1, \langle \nu(\cdot, \ell), \ell(\cdot) \rangle)$, and observe that $\Phi(M) = (L^1, L^2)$, if these quantities are defined as empirical neighbourhood, colour, and pair measures of a coloured graph. Note that while the *first* component of Φ is a continuous function, the *second* component is *discontinuous* in the weak topology.

To formulate the large deviations principle, we call a pair of measures $(\varpi, \nu) \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ *sub-consistent* if

$$\langle \nu(\cdot, \ell), \ell(\cdot) \rangle(a, b) \le \varpi(a, b), \quad \text{for all } a, b \in \mathcal{X}, \tag{2.1}$$

and *consistent* if equality holds in (2.1). Observe that, if ν is the empirical neighbourhood measure and ϖ the empirical pair measure of a coloured graph, (ϖ, ν) is consistent and both sides in (2.1) represent

$$\frac{1}{n} (1 + \mathbb{1}_{\{a=b\}}) \, \sharp\{\text{edges between vertices of colours } a \text{ and } b\}.$$

Recall the definition $H(\omega \| \mu) = \sum_{a \in \mathcal{X}} \omega(x) \log(\omega(x)/\mu(x))$ of the *relative entropy* of a probability measure $\omega \in \mathcal{M}(\mathcal{X})$ with respect to $\mu \in \mathcal{M}(\mathcal{X})$ and, for a measure $\varpi \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ and a measure $\omega \in \mathcal{M}(\mathcal{X})$, define

$$\mathfrak{H}_C(\varpi \| \omega) := H(\varpi \| C\omega \otimes \omega) + \|C\omega \otimes \omega\| - \|\varpi\|,$$

where the measure $C\omega \otimes \omega \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ is defined by $C\omega \otimes \omega(a,b) = C(a,b)\omega(a)\omega(b)$ for $a, b \in \mathcal{X}$. It is not hard to see that $\mathfrak{H}_C(\varpi \| \omega) \geq 0$. For every $(\varpi, \nu) \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ define a probability measure $Q = Q[\varpi, \nu]$ on $\mathcal{X} \times \mathcal{N}(\mathcal{X})$ by

$$Q(a\,,\ell) := \nu_1(a) \prod_{b \in \mathcal{X}} \frac{e^{-\left[\varpi(a,b)/\nu_1(a)\right]} \left[\varpi(a,b)/\nu_1(a)\right]^{\ell(b)}}{\ell(b)!}, \quad \text{for } a \in \mathcal{X}, \, \ell \in \mathcal{N}(\mathcal{X}).$$

We have now set the stage to state our principal theorem, the large deviation principle for the empirical pair measure and the empirical neighbourhood measure.

Theorem 2.1. *Suppose that X is a coloured random graph with colour law μ and connection probabilities $p_n \colon \mathcal{X} \times \mathcal{X} \to [0,1]$ satisfying $np_n(a,b) \to C(a,b)$ for some symmetric function $C \colon \mathcal{X} \times \mathcal{X} \to [0, \infty)$. Then, as $n \to \infty$, the pair $(L^2,\, M)$ satisfies a large deviation principle in $\tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with good rate function*

$$J(\varpi, \nu) = \begin{cases} H(\nu \,\|\, Q) + H(\nu_1 \,\|\, \mu) + \frac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \nu_1) & \text{if } (\varpi, \nu) \text{ sub-consistent,} \\ \infty & \text{otherwise.} \end{cases}$$

Remark 1 The rate function can be interpreted as follows: $J(\varpi\,,\nu)$ represents the cost of obtaining an empirical pair measure ϖ and an empirical neighbourhood measure ν. This cost is divided into three sub-costs:

  (i) $H(\nu_1 \,\|\, \mu)$ represents the cost of obtaining the empirical colour measure ν_1, this cost is non-negative and vanishes iff $\nu_1 = \mu$,
 (ii) $\frac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \nu_1)$ represent the cost of obtaining an empirical pair measure ϖ if the empirical colour measure is ν_1, again this cost is nonnegative and vanishes iff $\varpi = C\,\nu_1 \otimes \nu_1$,
(iii) $H(\nu \,\|\, Q)$ represents the cost of obtaining an empirical neighbourhood measure ν if the empirical colour measure is ν_1 and the empirical pair measure is ϖ, this cost is nonnegative and vanishes iff $\nu = Q$.

Consequently, $J(\varpi\,,\nu)$ is nonnegative and vanishes iff $\varpi = C\,\nu_1 \otimes \nu_1$ and

$$\nu(a,\ell) = \mu(a) \prod_{b \in \mathcal{X}} e^{-(C(a,b)\mu(b))} \frac{(C(a,b)\mu(b))^{\ell(b)}}{\ell(b)!}, \qquad \text{for all } (a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X}).$$

This is the law of a pair (a,ℓ) where a is distributed according to μ and, given the value of a, the random variables $\ell(b)$ are independently Poisson distributed with parameter $C(a,b)\mu(b)$.

Remark 2 Our large deviation principle implies individual large deviation principles for the measures L^2 and M by contraction, see [DZ98, Theorem 4.2.1]. Note that, by the discontinuity of Φ, the functional relationship $L^2 = \Phi_2(M)$ may break down in the limit, and hence the rate function may be finite on pairs which are not consistent.

As usual, the *degree distribution* $D \in \mathcal{M}(\mathbb{N} \cup \{0\})$ of a graph with empirical neighbourhood measure M is defined by

$$D(k) = \sum_{a \in \mathcal{X}} \sum_{\ell \in \mathcal{N}(\mathcal{X})} \delta_k\big(\textstyle\sum_{b \in \mathcal{X}} \ell(b)\big)\, M(a,\ell), \qquad \text{for } k \in \mathbb{N} \cup \{0\}.$$

As the degree distribution D is a continuous function of M, Theorem 2.1 and the contraction principle imply a large deviation principle for D. For a classical Erdős-Rényi graph the rate function takes on a particularly simple form.

Corollary 2.2. *Suppose D is the degree sequence of an Erdős-Rényi graph with connection probability $p_n \in [0,1]$ satisfying $np_n \to c \in (0, \infty)$. Then D satisfies a large deviation principle, as $n \to \infty$, in the space $\mathcal{M}(\mathbb{N} \cup \{0\})$ with good rate function*

$$\delta(d) = \begin{cases} \frac{1}{2} \langle d \rangle \log\left(\frac{\langle d \rangle}{c}\right) - \frac{1}{2}\langle d \rangle + \frac{c}{2} + H(d \,\|\, q_{\langle d \rangle}), & \text{if } \langle d \rangle > c, \\ \frac{1}{2} x \log\left(\frac{x}{c}\right) - \frac{1}{2}x + \frac{c}{2} + H(d \,\|\, q_x), & \text{if } \langle d \rangle \leq c, \end{cases} \tag{2.2}$$

where $x = x(d)$ is the unique solution of

$$x = ce^{-2\left(1 - \frac{\langle d \rangle}{x}\right)},$$

where q_λ is a Poisson distribution with parameter λ and $\langle d \rangle := \sum_{m=0}^{\infty} m d(m)$.

Remark 3 On probability measures d with mean c the rate simplifies to the relative entropy of d with respect to the Poisson distribution of the same mean. In [BGL02, Theorem 7.1] a large deviation principle for the degree sequence is formulated for this situation, albeit with a rather implicitly defined rate function. Moreover, the proof given there contains a serious gap: The exponential equivalence stated in [BGL02, Lemma 7.2] is not proved there and we conjecture that it does not hold.

We now state the two large deviation results, Theorems 2.3 and 2.5, which are the main ingredients for our proof of Theorem 2.1, but are also of independent interest. The first of these is a joint large deviation principle for the empirical colour measure L^1 and the empirical pair measure L^2, the second a large deviation principle for the empirical neighbourhood measure M given L^1 and L^2.

Theorem 2.3. *Suppose that X is a coloured random graph with colour law μ and connection probabilities satisfying $np_n(a,b) \to C(a,b)$ for some symmetric function $C \colon \mathcal{X} \times \mathcal{X} \to [0, \infty)$. Then, as $n \to \infty$, the pair (L^1, L^2) satisfies a large deviation principle in $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ with good rate function*

$$I(\omega, \varpi) = H(\omega \,\|\, \mu) + \tfrac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega). \tag{2.3}$$

Example 1 We look at the Erdős-Rényi graph with connection probabilities p_n satisfying $np_n \to c \in (0, \infty)$ and study the random partition function for the *Ising model* on the graph, which is defined as

$$Z(\beta) := \sum_{\eta \in \{-1, +1\}^V} \exp\left(\beta \sum_{(u,v) \in E} \eta(u)\eta(v)\right) \text{ for the inverse temperature } \beta > 0.$$

Denoting by E expectation with respect to the graph, we note that

$$\mathsf{E}Z(\beta) = 2^n \mathbb{E} \exp\left(n \tfrac{\beta}{2} \int xy \, L^2(dx \, dy)\right),$$

where $\mathbb{E}$ is expectation with respect to the graph randomly coloured using independent colours chosen uniformly from $\{-1, 1\}$. Then Varadhan's lemma, see e.g. [dH00, III.3], Theorem 2.3, gives

$$\lim_{n \to \infty} \tfrac{1}{n} \log \mathsf{E}Z(\beta)$$

$$= \log 2 + \sup\left\{\tfrac{\beta}{2} \int xy \, \varpi(dx \, dy) - I(\omega, \varpi) : \omega \in \mathcal{M}(\mathcal{X}), \varpi \in \mathcal{M}_*(\{-1, 1\} \times \{-1, 1\})\right\} \tag{2.4}$$

$$= \sup\left\{\tfrac{\beta}{2}\left(\varpi(\Delta) - \varpi(\Delta^c)\right) - x\log(x) - (1-x)\log(1-x) - \tfrac{1}{2}\left(H(\varpi \,\|\, \omega_x) + c - \|\varpi\|\right)\right\},$$

where Δ is the diagonal in $\{-1, 1\} \times \{-1, 1\}$, and the supremum is over all $x \in [0,1]$ and $\varpi \in \mathcal{M}_*(\{-1, 1\} \times \{-1, 1\})$, and the measure $\omega_x \in \tilde{\mathcal{M}}_*(\{-1, 1\} \times \{-1, 1\})$ is defined by

$$\omega_x(i,j) = cx^{(2+i+j)/2}(1-x)^{(2-i-j)/2} \text{ for } i, j \in \{-1, 1\}.$$

Note that the last expression in (2.4) is an optimisation problem in only four real variables.

Corollary 2.4. *Suppose that X is a coloured random graph with colour law μ and connection probabilities satisfying $np_n(a,b) \to C(a,b)$ for some symmetric function $C \colon \mathcal{X} \times \mathcal{X} \to [0, \infty)$. Then, as $n \to \infty$, the number of edges per vertex $|E|/n$ satisfies a large deviation principle in $[0, \infty)$ with good rate function*

$$\zeta(x) = x \log x - x + \inf_{y > 0} \left\{ \psi(y) - x \log(\tfrac{1}{2} y) + \tfrac{1}{2} y \right\},$$

where $\psi(y) = \inf H(\omega \,\|\, \mu)$ over all probability vectors ω with $\omega^T C \omega = y$.

Remark 4 In the Erdős-Rényi case $C(a,b) = c$ one obtains $\psi(y) = 0$ for $y = c$, and $\psi(y) = \infty$ otherwise. Hence $\zeta(x) = x \log x - x - x \log(\frac{c}{2}) + \frac{c}{2}$, which is the Cramér rate function for the Poisson distribution with parameter $\frac{c}{2}$. In [BP03] a large deviation principle for $|E|/n^2$ is proved for coloured random graphs with *fixed* connection probabilities.

For any $n \in \mathbb{N}$ we define

$$\mathcal{M}_n(\mathcal{X}) := \left\{ \omega \in \mathcal{M}(\mathcal{X}) \,:\, n\omega(a) \in \mathbb{N} \text{ for all } a \in \mathcal{X} \right\},$$

$$\tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X}) := \left\{ \varpi \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \,:\, \tfrac{n}{1 + \mathbb{1}_{\{a=b\}}} \varpi(a,b) \in \mathbb{N} \text{ for all } a, b \in \mathcal{X} \right\},$$

We write

$$\mathfrak{M}[\omega, \varpi] = \left\{ \nu \colon (\varpi, \nu) \text{ sub-consistent and } \nu_1 = \omega \right\} \subset \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})).$$

Theorem 2.5. *Suppose the sequence $(\omega_n, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$ converges to a limit $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$. Let X be a coloured random graph with n vertices conditioned on the event $\{\Phi(M) = (\omega_n, \varpi_n)\}$. Then, as $n \to \infty$, the empirical neighbourhood measure M of X satisfies a large deviation principle in the space $\mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with good rate function*

$$\tilde{J}_{(\omega,\varpi)}(\nu) = \begin{cases} H(\nu \,\|\, Q) & \text{if } \nu \in \mathfrak{M}[\omega, \varpi] \\ \infty & \text{otherwise.} \end{cases} \tag{2.5}$$

In the remainder of the paper we give the proofs of the results set out so far. Section 3 is devoted to the proof of Theorem 2.3, which uses a change-of-measure technique. By contrast, the proof of Theorem 2.5, carried out in Section 4, is based on combinatorial arguments combined with a fairly sophisticated probabilistic approximation technique. In Section 5 we combine these results to obtain our main result, Theorem 2.1, using the setup and result of Biggins [Bi04] to 'mix' the large deviation principles. The paper concludes with the proofs of Corollaries 2.2 and 2.4, which are given in Section 6.

3. Proof of Theorem 2.3 by change-of-measure

Given a function $\tilde{f} \colon \mathcal{X} \to \mathbb{R}$ and a symmetric function $\tilde{g} \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ we define the constant $U_{\tilde{f}}$ by

$$U_{\tilde{f}} = \log \sum_{a \in \mathcal{X}} e^{\tilde{f}(a)} \mu(a),$$

and the function $\tilde{h}_n \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde{h}_n(a,b) = \log \left[\left(1 - p_n(a,b) + p_n(a,b) e^{\tilde{g}(a,b)} \right)^{-n} \right], \tag{3.1}$$

for $a, b \in \mathcal{X}$. We use $\tilde{f}$ and $\tilde{g}$ to define (for sufficiently large n) a new coloured random graph as follows:

- To the n labelled vertices in V we assign colours from $\mathcal{X}$ independently and identically according to the colour law $\widetilde{\mu}$ defined by

$$\widetilde{\mu}(a) = e^{\tilde{f}(a) - U_{\tilde{f}}} \mu(a).$$

- Given any two vertices $u, v \in V$, with u carrying colour a and v carrying colour b, we connect vertex u to vertex v with probability

$$\tilde{p}_n(a,b) = \frac{p_n(a,b)e^{\tilde{g}(a,b)}}{1 - p_n(a,b) + p_n(a,b)e^{\tilde{g}(a,b)}}.$$

We denote the transformed law by $\tilde{\mathbb{P}}$. We observe that $\tilde{\mu}$ is a probability measure and that $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$ as, for any coloured graph $X = ((X(v) \colon v \in V), E)$,

$$
\begin{aligned}
\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) &= \prod_{u \in V} \frac{\tilde{\mu}(X(u))}{\mu(X(u))} \prod_{(u,v) \in E} \frac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \prod_{(u,v) \notin E} \frac{1-\tilde{p}_n(X(u),X(v))}{1-p_n(X(u),X(v))} \\
&= \prod_{u \in V} \frac{\tilde{\mu}(X(u))}{\mu(X(u))} \prod_{(u,v) \in E} \frac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \times \frac{n-np_n(X(u),X(v))}{n-n\tilde{p}_n(X(u),X(v))} \prod_{(u,v) \in \mathcal{E}} \frac{n-n\tilde{p}_n(X(u),X(v))}{n-np_n(X(u),X(v))} \\
&= \prod_{u \in V} e^{\tilde{f}(X(u))-U_{\tilde{f}}} \prod_{(u,v) \in E} e^{\tilde{g}(X(u),X(v))} \prod_{(u,v) \in \mathcal{E}} e^{\frac{1}{n}\tilde{h}_n(X(u),X(v))} \\
&= \exp\left(n\langle L^1, \tilde{f} - U_{\tilde{f}} \rangle + n\langle \tfrac{1}{2}L^2, \tilde{g} \rangle + n\langle \tfrac{1}{2}L^1 \otimes L^1, \tilde{h}_n \rangle - \langle \tfrac{1}{2}L^1_{\Delta}, \tilde{h}_n \rangle \right), \qquad (3.2)
\end{aligned}
$$

where $L^1_{\Delta} = \frac{1}{n}\sum_{u \in V} \delta_{(X(u),X(u))}$ and we denote $\langle g, \varpi \rangle := \sum_{a,b \in \mathcal{X}} g(a,b)\varpi(a,b)$ for $\varpi \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$, and $\langle f, \omega \rangle := \sum_{a \in \mathcal{X}} f(a)\omega(a)$ for $\omega \in \mathcal{M}(\mathcal{X})$. The following lemmas will be useful later.

Lemma 3.1 (Euler's lemma). *If $np_n(a,b) \to C(a,b)$ for every $a, b \in \mathcal{X}$, then*

$$\lim_{n \to \infty} \left[1 + \alpha p_n(a,b)\right]^n = e^{\alpha C(a,b)}, \text{ for all } a, b \in \mathcal{X} \text{ and } \alpha \in \mathbb{R}. \qquad (3.3)$$

Proof.   Observe that, for any $\varepsilon > 0$ and for large n we have

$$\left[1 + \tfrac{\alpha C(a,b)-\varepsilon}{n}\right]^n \le \left[1 + \alpha p_n(a,b)\right]^n \le \left[1 + \tfrac{\alpha C(a,b)+\varepsilon}{n}\right]^n,$$

by the pointwise convergence. Hence by the sandwich theorem and Euler's formula we get (3.3).   ∎

Lemma 3.2 (Exponential tightness). *For every $\theta > 0$ there exists $N \in \mathbb{N}$ such that*

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\Big\{|E| > nN\Big\} \le -\theta.$$

Proof.   Let $c > \max_{a,b \in \mathcal{X}} C(a,b) > 0$. By a simple coupling argument we can define, for all sufficiently large n, a new coloured random graph $\tilde{X}$ with colour law μ and connection probability $\frac{c}{n}$, such that any edge present in X is also present in $\tilde{X}$. Let $|\tilde{E}|$ be the number of edges of $\tilde{X}$. Using Chebyshev's inequality, the binomial formula, and Lemma 3.1, we have that

$$
\begin{aligned}
\mathbb{P}\Big\{|\tilde{E}| \ge nl\Big\} &\le e^{-nl}\mathbb{E}\{e^{|\tilde{E}|}\} = e^{-nl} \sum_{k=0}^{\frac{n(n-1)}{2}} e^k \binom{n(n-1)/2}{k} \left(\frac{c}{n}\right)^k \left(1 - \frac{c}{n}\right)^{n(n-1)/2-k} \\
&= e^{-nl}\left(1 - \frac{c}{n} + e\frac{c}{n}\right)^{n(n-1)/2} \le e^{-nl}e^{nc(e-1+o(1))}.
\end{aligned}
$$

Now given $\theta > 0$ choose $N \in \mathbb{N}$ such that $N > \theta + c(e-1)$ and observe that, for sufficiently large n,

$$\mathbb{P}\{|E| \ge nN\} \le \mathbb{P}\{|\tilde{E}| \ge nN\} \le e^{-n\theta},$$

which implies the statement.   ∎

3.1 Proof of the upper bound in Theorem 2.3

We denote by $\mathcal{C}_1$ the space of functions on $\mathcal{X}$ and by $\mathcal{C}_2$ the space of symmetric functions on $\mathcal{X} \times \mathcal{X}$, and define

$$\hat{I}(\omega, \varpi) = \sup_{\substack{f \in \mathcal{C}_1 \\ g \in \mathcal{C}_2}} \left\{ \sum_{a \in \mathcal{X}} (f(a) - U_f)\omega(a) + \tfrac{1}{2} \sum_{a,b \in \mathcal{X}} g(a,b)\varpi(a,b) + \tfrac{1}{2} \sum_{a,b \in \mathcal{X}} (1 - e^{g(a,b)})C(a,b)\omega(a)\omega(b) \right\}$$

for $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$.

Lemma 3.3. *For each closed set $F \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$,*

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{ (L^1, L^2) \in F \big\} \leq - \inf_{(\omega, \varpi) \in F} \hat{I}(\omega, \varpi).$$

Proof.   Fix $\tilde{f} \in \mathcal{C}_1$ and $\tilde{g} \in \mathcal{C}_2$. Define $\tilde{\beta} \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by $\tilde{\beta}(a,b) = (1 - e^{\tilde{g}(a,b)})C(a,b)$. Observe that, by Lemma 3.1, $\tilde{\beta}(a,b) = \lim_{n \to \infty} \tilde{h}_n(a,b)$ for all $a, b \in \mathcal{X}$, recalling the definition of $\tilde{h}_n$ from (3.1). Hence, by (3.2), for sufficiently large n,

$$e^{\max_{a \in \mathcal{X}} |\beta(a,a)|} \geq \int e^{\langle \frac{1}{2} L_\Delta^1, \tilde{h}_n \rangle} d\tilde{\mathbb{P}} = \mathbb{E}\left\{ e^{n\langle L^1, \tilde{f} - U_{\tilde{f}} \rangle + n\langle \frac{1}{2} L^2, \tilde{g} \rangle + n\langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n \rangle} \right\},$$

where $L_\Delta^1 = \tfrac{1}{n} \sum_{u \in V} \delta_{(X(u), X(u))}$ and therefore,

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{E}\left\{ e^{n\langle L^1, \tilde{f} - U_{\tilde{f}} \rangle + n\langle \frac{1}{2} L^2, \tilde{g} \rangle + n\langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n \rangle} \right\} \leq 0. \tag{3.4}$$

We now fix $\varepsilon > 0$, and let $\hat{I}_\varepsilon(\omega, \varpi) = \min\{\hat{I}(\omega, \varpi), \varepsilon^{-1}\} - \varepsilon$. Suppose that $(\omega, \varpi) \in F$ and observe that $\hat{I}(\omega, \varpi) > \hat{I}_\varepsilon(\omega, \varpi)$. Choose $\tilde{f} \in \mathcal{C}_1$ and $\tilde{g} \in \mathcal{C}_2$ such that

$$\langle \tilde{f} - U_{\tilde{f}}, \omega \rangle + \tfrac{1}{2} \langle \tilde{g}, \varpi \rangle + \tfrac{1}{2} \langle \tilde{\beta}, \omega \otimes \omega \rangle \geq \hat{I}_\varepsilon(\omega, \varpi).$$

As $\mathcal{X}$ is finite, there exist open neighbourhoods B_ϖ^2 and B_ω^1 of ϖ, ω such that

$$\inf_{\substack{\tilde{\omega} \in B_\omega^1 \\ \tilde{\varpi} \in B_\varpi^2}} \left\{ \langle \tilde{f} - U_{\tilde{f}}, \tilde{\omega} \rangle + \tfrac{1}{2} \langle \tilde{g}, \tilde{\varpi} \rangle + \tfrac{1}{2} \langle \tilde{\beta}, \tilde{\omega} \otimes \tilde{\omega} \rangle \right\} \geq \hat{I}_\varepsilon(\omega, \varpi) - \varepsilon.$$

Using Chebyshev's inequality and (3.4) we have that

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{ (L^1, L^2) \in B_\omega^1 \times B_\varpi^2 \big\}$$

$$\leq \limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{E}\left\{ e^{n\langle L^1, \tilde{f} - U_{\tilde{f}} \rangle + n\langle \frac{1}{2} L^2, \tilde{g} \rangle + n\langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n \rangle} \right\} - \hat{I}_\varepsilon(\omega, \varpi) + \varepsilon \tag{3.5}$$

$$\leq -\hat{I}_\varepsilon(\omega, \varpi) + \varepsilon.$$

Now we use Lemma 3.2 with $M = \varepsilon^{-1}$, to choose $N(\varepsilon) \in \mathbb{N}$ such that

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\left\{ |E| > nN(\varepsilon) \right\} \leq -\varepsilon^{-1}. \tag{3.6}$$

For this $N(\varepsilon)$, define the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)} = \left\{ (\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) : \|\varpi\| \leq 2N(\varepsilon) \right\},$$

and recall that $\|L^2\| = 2\frac{|E|}{n}$. The set $K_{N(\varepsilon)} \cap F$ is compact and therefore may be covered by finitely many sets $B^1_{\omega_r} \times B^2_{\varpi_r}, r = 1, \ldots, m$ with $(\omega_r, \varpi_r) \in F$ for $r = 1, \ldots, m$. Consequently,

$$\mathbb{P}\big\{(L^1, L^2) \in F\big\} \leq \sum_{r=1}^{m} \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} + \mathbb{P}\big\{(L^1, L^2) \notin K_{N(\varepsilon)}\big\}.$$

We may now use (3.5) and (3.6) to obtain, for all sufficiently small $\varepsilon > 0$,

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in F\big\} \leq \max_{r=1}^{m} \limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\}$$

$$\leq -\inf_{(\omega, \varpi) \in F} \hat{I}_\varepsilon(\omega, \varpi) + \varepsilon.$$

Taking $\varepsilon \downarrow 0$ we have the desired statement. ∎

Next, we express the rate function in term of relative entropies, see for example [DZ98, (2.15)], and consequently show that it is a good rate function. Recall the definition of the function I from Theorem 2.3.

Lemma 3.4.
   (i) $\hat{I}(\omega, \varpi) = I(\omega, \varpi)$, for any $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$,
   (ii) I is a good rate function and
   (iii) $\mathfrak{H}_C(\varpi \,\|\, \omega) \geq 0$ with equality if and only if $\varpi = C\omega \otimes \omega$.

Proof.   (i) Suppose that $\varpi \not\ll C\omega \otimes \omega$. Then, there exists $a_0, b_0 \in \mathcal{X}$ with $C\omega \otimes \omega(a_0, b_0) = 0$ and $\varpi(a_0, b_0) > 0$. Define $\hat{g} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\hat{g}(a, b) = \log\big[K(\mathbb{1}_{(a_0, b_0)}(a, b) + \mathbb{1}_{(b_0, a_0)}(a, b)) + 1\big], \quad \text{for } a, b \in \mathcal{X} \text{ and } K > 0.$$

For this choice of $\hat{g}$ and $f = 0$ we have

$$\sum_{a \in \mathcal{X}} \big(f(a) - U_f\big)\omega(a) + \sum_{a,b \in \mathcal{X}} \tfrac{1}{2}\hat{g}(a, b)\varpi(a, b) + \sum_{a,b \in \mathcal{X}} \tfrac{1}{2}(1 - e^{\hat{g}(a,b)})C(a, b)\omega(a)\omega(b)$$

$$= \tfrac{1}{2}\log(K+1)\varpi(a_0, b_0) \to \infty, \qquad \text{for } K \uparrow \infty.$$

Now suppose that $\varpi \ll C\omega \otimes \omega$. We have

$$\hat{I}(\omega, \varpi) = \sup_{f \in \mathcal{C}_1} \Big\{ \sum_{a \in \mathcal{X}} \Big(f(a) - \log \sum_{a \in \mathcal{X}} e^{f(a)}\mu(a)\Big)\omega(a) \Big\}$$

$$+ \tfrac{1}{2} \sum_{a,b \in \mathcal{X}} C(a, b)\omega(a)\omega(b) + \tfrac{1}{2} \sup_{g \in \mathcal{C}_2} \Big\{ \sum_{a,b \in \mathcal{X}} g(a, b)\varpi(a, b) - \sum_{a,b \in \mathcal{X}} e^{g(a,b)}C(a, b)\omega(a)\omega(b) \Big\}.$$

By the variational characterization of entropy, the first term equals $H(\omega \,\|\, \mu)$. By the substitution $h = e^g \frac{C\omega \otimes \omega}{\varpi}$ the last term equals

$$\sup_{\substack{h \in \mathcal{C}_2 \\ h \geq 0}} \sum_{a,b \in \mathcal{X}} \Big[\log\Big(h(a, b)\frac{\varpi(a, b)}{C(a, b)\omega(a)\omega(b)}\Big) - h(a, b)\Big]\varpi(a, b)$$

$$= \sup_{\substack{h \in \mathcal{C}_2 \\ h \geq 0}} \sum_{a,b \in \mathcal{X}} \big(\log h(a, b) - h(a, b)\big)\varpi(a, b) + \sum_{a,b \in \mathcal{X}} \log\Big(\frac{\varpi(a, b)}{C(a, b)\omega(a)\omega(b)}\Big)\varpi(a, b)$$

$$= -\|\varpi\| + H(\varpi \,\|\, C\omega \otimes \omega),$$

where we have used $\sup_{x>0} \log x - x = -1$ in the last step. This yields that $\hat{I}(\omega, \varpi) = I(\omega, \varpi)$.

(ii) Recall from (2.3) and the definition of $\mathfrak{H}_C$ that $I(\omega, \varpi) = H(\omega \| \mu) + \frac{1}{2} H(\varpi \| C\omega \otimes \omega) + \frac{1}{2} \|C\omega \otimes \omega\| - \frac{1}{2} \|\varpi\|$. All summands are continuous in ω, ϖ and thus I is a rate function. Moreover, for all $\alpha < \infty$, the level sets $\{I \le \alpha\}$ are contained in the bounded set $\{(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \colon \mathfrak{H}_C(\varpi \| \omega) \le \alpha\}$ and are therefore compact. Consequently, I is a good rate function.

(iii) Consider the nonnegative function $\xi(x) = x \log x - x + 1$, for $x > 0$, $\xi(0) = 1$, which has its only root in $x = 1$. Note that

$$\mathfrak{H}_C(\varpi \| \omega) = \begin{cases} \int \xi \circ g \, dC\omega \otimes \omega & \text{if } g := \frac{d\varpi}{dC\omega \otimes \omega} \ge 0 \text{ exists} \\ \infty & \text{otherwise.} \end{cases} \tag{3.7}$$

Hence $\mathfrak{H}_C(\varpi \| \omega) \ge 0$, and, if $\varpi = C\omega \otimes \omega$, then $\xi(\frac{d\varpi}{dC\omega \otimes \omega}) = \xi(1) = 0$ and so $\mathfrak{H}_C(C\omega \otimes \omega \| \omega) = 0$. Conversely, if $\mathfrak{H}_C(\varpi \| \omega) = 0$, then $\varpi(a, b) > 0$ implies $C\omega \otimes \omega(a, b) > 0$, which then implies $\xi \circ g(a, b) = 0$ and further $g(a, b) = 1$. Hence $\varpi = C\omega \otimes \omega$, which completes the proof of (iii). ∎

3.2 Proof of the lower bound in Theorem 2.3

Lemma 3.5. *For every open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$,*

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\Big\{(L^1, L^2) \in O\Big\} \ge - \inf_{(\omega, \varpi) \in O} I(\omega, \varpi).$$

Proof.   Suppose $(\omega, \varpi) \in O$, with $\varpi \ll C\omega \otimes \omega$. Define $\tilde{f}_\omega \colon \mathcal{X} \to \mathbb{R}$ by

$$\tilde{f}_\omega(a) = \begin{cases} \log \frac{\omega(a)}{\mu(a)}, & \text{if } \omega(a) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

and $\tilde{g}_\varpi \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde{g}_\varpi(a, b) = \begin{cases} \log \frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}, & \text{if } \varpi(a, b) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

In addition, we let $\tilde{\beta}_\varpi(a, b) = C(a, b)(1 - e^{\tilde{g}_\varpi(a,b)})$ and note that $\tilde{\beta}_\varpi(a, b) = \lim_{n \to \infty} \tilde{h}_{\varpi, n}(a, b)$, for all $a, b \in \mathcal{X}$ where

$$\tilde{h}_{\varpi, n}(a, b) = \log \left[\left(1 - p_n(a, b) + p_n(a, b) e^{\tilde{g}_\varpi(a,b)}\right)^{-n} \right].$$

Choose B^1_ω, B^2_ϖ open neighbourhoods of ω, ϖ, such that $B^1_\omega \times B^2_\varpi \subset O$ and for all $(\tilde{\omega}, \tilde{\varpi}) \in B^1_\omega \times B^2_\varpi$

$$\langle \tilde{f}_\omega, \omega \rangle + \tfrac{1}{2} \langle \tilde{g}_\varpi, \varpi \rangle + \tfrac{1}{2} \langle \tilde{\beta}_\varpi, \omega \otimes \omega \rangle - \varepsilon \le \langle \tilde{f}_\omega, \tilde{\omega} \rangle + \tfrac{1}{2} \langle \tilde{g}_\varpi, \tilde{\varpi} \rangle + \tfrac{1}{2} \langle \tilde{\beta}_\varpi, \tilde{\omega} \otimes \tilde{\omega} \rangle.$$

We now use $\tilde{\mathbb{P}}$, the probability measure obtained by transforming $\mathbb{P}$ using the functions $\tilde{f}_\omega, \tilde{g}_\varpi$. Note that the colour law in the transformed measure is now ω, and the connection probabilities $\tilde{p}_n(a, b)$ satisfy $n \tilde{p}_n(a, b) \to \varpi(a, b)/(\omega(a)\omega(b)) =: \tilde{C}(a, b)$, as $n \to \infty$. Using (3.2), we obtain

$$\mathbb{P}\Big\{(L^1, L^2) \in O\Big\} \ge \tilde{\mathbb{E}}\Big\{ \tfrac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(X) \mathbb{1}_{\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\}} \Big\}$$

$$= \tilde{\mathbb{E}}\Big\{ \prod_{u \in V} e^{-\tilde{f}_\omega(X(u))} \prod_{(u,v) \in E} e^{-\tilde{g}_\varpi(X(u), X(v))} \prod_{(u,v) \in \mathcal{E}} e^{-\frac{1}{n} \tilde{h}_{\varpi, n}(X(u), X(v))} \mathbb{1}_{\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\}} \Big\}$$

$$= \tilde{\mathbb{E}}\Big\{ e^{-n\langle L^1, \tilde{f}_\omega \rangle - n\frac{1}{2}\langle L^2, \tilde{g}_\varpi \rangle - n\frac{1}{2}\langle L^1 \otimes L^1, \tilde{g}_\varpi \rangle + \frac{1}{2}\langle L^1_\Delta, \tilde{h}_{\varpi, n} \rangle} \times \mathbb{1}_{\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\}} \Big\}$$

$$\ge \exp\Big(-n\langle \tilde{f}_\omega, \omega \rangle - n\tfrac{1}{2}\langle \tilde{g}_\varpi, \varpi \rangle - n\tfrac{1}{2}\langle \tilde{\beta}_\varpi, \omega \otimes \omega \rangle + m - n\varepsilon \Big) \times \tilde{\mathbb{P}}\Big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\Big\},$$

where $m := 0 \wedge \min_{a \in \mathcal{X}} \tilde{\beta}(a, a)$. Therefore, by (3.3), we have

$$\liminf_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\Big\{(L^1, L^2) \in O\Big\}$$
$$\geq -\langle \tilde{f}_\omega, \omega \rangle - \tfrac{1}{2} \langle \tilde{g}_\varpi, \varpi \rangle - \tfrac{1}{2} \langle \tilde{\beta}_\varpi, \omega \otimes \omega \rangle - \varepsilon + \liminf_{n \to \infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\Big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\Big\}.$$

The result follows once we prove that

$$\liminf_{n \to \infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\Big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\Big\} = 0. \tag{3.8}$$

We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to prove (3.8). Then we obtain

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\{(L^1, L^2) \in (B^1_\omega \times B^2_\varpi)^c\} \leq - \inf_{(\tilde{\omega}, \tilde{\varpi}) \in \tilde{F}} \tilde{I}(\tilde{\omega}, \tilde{\varpi}),$$

where $\tilde{F} = (B^1_\omega \times B^2_\varpi)^c$ and $\tilde{I}(\tilde{\omega}, \tilde{\varpi}) := H(\tilde{\omega} \,\|\, \omega) + \tfrac{1}{2}\mathfrak{H}_{\tilde{C}}(\tilde{\varpi} \,\|\, \tilde{\omega})$. It therefore suffices to show that the infimum is positive. Suppose for contradiction that there exists a sequence $(\tilde{\omega}_n, \tilde{\varpi}_n) \in \tilde{F}$ with $\tilde{I}(\tilde{\omega}_n, \tilde{\varpi}_n) \downarrow 0$. Then, because $\tilde{I}$ is a good rate function and its level sets are compact, and by lower semicontinuity of the mapping $(\tilde{\omega}, \tilde{\varpi}) \mapsto \tilde{I}(\tilde{\omega}, \tilde{\varpi})$, we can construct a limit point $(\tilde{\omega}, \tilde{\varpi}) \in \tilde{F}$ with $\tilde{I}(\tilde{\omega}, \tilde{\varpi}) = 0$. By Lemma 3.4 this implies $H(\tilde{\omega} \,\|\, \omega) = 0$ and $\mathfrak{H}_C(\tilde{\varpi} \,\|\, \tilde{\omega}) = 0$, hence $\tilde{\omega} = \omega$, and $\tilde{\varpi} = \tilde{C}\tilde{\omega} \otimes \tilde{\omega} = \varpi$ contradicting $(\tilde{\omega}, \tilde{\varpi}) \in \tilde{F}$. ∎

4. Proof of Theorem 2.5 by the method of types

Throughout the proof we may assume that $\omega(a) > 0$ for all $a \in \mathcal{X}$. It is easy to see that the law of the randomly coloured graph conditioned to have empirical colour measure ω_n and empirical pair measure ϖ_n,

$$\mathbb{P}_{(\omega_n, \varpi_n)} = \mathbb{P}\{\,\cdot\, \,|\, \Phi(M) = (\omega_n, \varpi_n)\},$$

can be described in the following manner:

- Assign colours to the vertices by sampling without replacement from the collection of n colours, which contains any colour $a \in \mathcal{X}$ exactly $n\omega_n(a)$ times;
- for every unordered pair $\{a, b\}$ of colours create exactly $n(a, b)$ edges by sampling without replacement from the pool of possible edges connecting vertices of colour a and b, where

$$n(a, b) := \begin{cases} n\,\varpi_n(a, b) & \text{if } a \neq b \\ \tfrac{n}{2}\,\varpi_n(a, b) & \text{if } a = b. \end{cases}$$

For our proof it is convenient to introduce a numbering system, which specifies, for each $\{a, b\}$, the order in which edges are drawn in the second step. More precisely, the edge-number k is attached to both vertices connecting the kth edge. Note that the total number of edge-numbers attached to every vertex corresponds to the degree of the vertex in the graph. All permitted numberings are equally probable, with the total number of possible numberings (given the coloured graph) being

$$\prod_{\{a,b\}} n(a, b)! \ .$$

Denote by $V(a)$ the set of vertices with colour a, and let $Y_j^{\{a,b\}}$ be the jth edge drawn in the process of connecting vertices of colours $\{a, b\}$. Let $\mathcal{A}_n(\omega_n, \varpi_n)$ be the set of all possible configurations

$$\tilde{X} = \Big((V(a) : a \in \mathcal{X}); \big(Y_k^{\{a,b\}} : k = 1, \ldots, n(a, b); \{a, b\} \subset \mathcal{X}\big)\Big),$$

and let $\mathcal{B}_n(\omega_n, \varpi_n)$ be the set of all coloured graphs x with $L^1(x) = \omega_n$ and $L^2(x) = \varpi_n$. Define $\Psi \colon \mathcal{A}_n(\omega_n, \varpi_n) \to \mathcal{B}_n(\omega_n, \varpi_n)$ as the canonical mapping which associates the coloured graph to any configuration, i.e. 'forgets' the numbering of the edges. Finally, define

$$\mathcal{K}^{(n)}(\omega_n, \varpi_n) := \left\{ M(x) \text{ for some } x \in \mathcal{B}_n(\omega_n, \varpi_n) \right\}$$

to be the set of all empirical neighbourhood measures $M(x)$ arising from coloured graphs x with n vertices with $\Phi(M(x)) = (\omega_n, \varpi_n)$.

In our proofs we use the following form of Stirling's formula, see [Fel67, page 54]: For all $n \in \mathbb{N}$,

$$n^n \, e^{-n} \le n! \le (2\pi n)^{\frac{1}{2}} \, n^n \, e^{-n+1/(12n)} \, .$$

4.1 A bound on the number of empirical neighbourhood measures

In this section we provide an upper bound on the number of measures in $\mathcal{K}^{(n)}(\omega_n, \varpi_n)$. We write m for the number of elements in $\mathcal{X}$.

Lemma 4.1. *There exists $\vartheta = \vartheta(m) > 0$ such that, if $\omega_n \in \mathcal{M}_n(\mathcal{X})$ and $\varpi_n \in \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$, then*

$$\sharp \mathcal{K}^{(n)}(\omega_n, \varpi_n) \le \exp \left[\vartheta \, (\log n) \, (n \|\varpi_n\|)^{\frac{2m-1}{2m}} \right]. \tag{4.1}$$

The proof is based on counting integer partitions of vectors. To fix some notation, let $\mathfrak{I}_m = \left(\mathbb{N} \cup \{0\} \right)^m$ be the collection of (nonnegative) integer vectors of length m. For any $\ell \in \mathfrak{I}_m$ we denote by $\|\ell\|$ its magnitude, i.e. the sum of its entries.

We introduce an ordering $\succeq$ on $\mathfrak{I}_m$ such that, for any vectors

$$\ell_1 = (\ell_1^{(1)}, \ldots, \ell_1^{(m)}) \text{ and } \ell_2 = (\ell_2^{(1)}, \ldots, \ell_2^{(m)}),$$

we write $\ell_1 \succeq \ell_2$ if either

    (i) $\|\ell_1\| > \|\ell_2\|$, *or*

    (ii) $\|\ell_1\| = \|\ell_2\|$ and there is $j \in \{1, \ldots, m\}$ with $\ell_1^{(k)} = \ell_2^{(k)}$, for all $k < j$, and $\ell_1^{(j)} > \ell_2^{(j)}$, *or*

    (iii) $\ell_1 = \ell_2$.

A collection $(\ell_1, \ldots, \ell_k)$ of elements in $\mathfrak{I}_m$ is an *integer partition* of the vector $\ell \in \mathfrak{I}_m$, if

$$\ell_1 \succeq \ldots \succeq \ell_k \neq 0 \qquad \text{and} \qquad \ell_1 + \ldots + \ell_k = \ell.$$

Any integer partition of a vector $\ell \in \mathfrak{I}_m$ induces an integer partition $\|\ell_1\|, \ldots, \|\ell_k\|$ of its magnitude $\|\ell\|$, which we call its *sum-partition*. We denote by $\mathcal{P}_m(\ell)$ the set of integer partitions of ℓ.

Lemma 4.2. *There exists $\vartheta = \vartheta(m) > 0$ such that, for any $\ell \in \mathfrak{I}_m$ of magnitude n,*

$$\sharp \mathcal{P}_m(\ell) \le \exp \left[\vartheta \, (\log n) \, n^{\frac{2m-1}{2m}} \right].$$

Proof.   Let $\ell \in \mathfrak{I}_m$ be a vector of magnitude n and $(\ell_1, \ldots, \ell_k)$ be an integer partition of ℓ. We rewrite the partition as $(\mathfrak{m}_{1,1}, \ldots, \mathfrak{m}_{1,k_1}; \mathfrak{m}_{2,1}, \ldots, \mathfrak{m}_{2,k_2}; \ldots; \mathfrak{m}_{r,1}, \ldots, \mathfrak{m}_{r,k_r})$ such that all vectors in the same block (indicated by the first subscript) have the same magnitude, which we denote $y_1, \ldots, y_r$, and such that $y_1 > \cdots > y_r > 0$. Note that for the block sizes we have $k_1 + \cdots + k_r = k$ and that $k_1 y_1 + \cdots + k_r y_r = n$.

For a moment, look at a fixed block $\mathfrak{m}_{j,1}, \ldots, \mathfrak{m}_{j,k_j}$. It is easy to see that the number of integer vectors of length m and magnitude y_j is given by

$$b(y_j, m) := \binom{y_j + m - 1}{m - 1} \le c(m) \, y_j^{m-1} \, .$$

Writing $\mathfrak{m}_{j,0}$ for the largest and $\mathfrak{m}_{j,k_j+1}$ for the smallest of these vectors in the ordering $\succcurlyeq$, we note that

$$p\colon \{0,\ldots,k_j+1\} \to \{\mathfrak{m} \in \mathfrak{I}_m\colon \|\mathfrak{m}\| = y_j\}, \qquad p(i) = \mathfrak{m}_{j,i},$$

is a non-increasing path of length k_j+2 into an ordered set of size $b(y_j,m)$, which connects the smallest to the largest element. The number of such paths is easily seen to be

$$\binom{b(y_j,m)+k_j}{k_j}.$$

Therefore, the number of integer partitions of ℓ with given sum-partition $(y_1,\overset{k_1}{\ldots},y_1,\ldots,y_r,\overset{k_r}{\ldots},y_r)$ is

$$\prod_{j=1}^{r} \binom{b(y_j,m)+k_j}{k_j} \le \max_{\substack{a_1,\ldots,a_r>0 \\ \sum a_j=n}} \prod_{j=1}^{r} \left\{ \max_{\substack{y,k\in\mathbb{N} \\ yk=a_j}} \binom{c(m)y^{m-1}+k}{k} \right\}.$$

To maximize the binomial coefficient over the set $yk = a_j$, we distinguish between the cases when (i) $a_j \le c(m)y^m$, (ii) $a_j > c(m)y^m$ and observe that

$$\binom{c(m)y^{m-1}+\frac{a_j}{y}}{\frac{a_j}{y}} \le \begin{cases} \binom{2c(m)y^{m-1}}{\frac{a_j}{y}} & \text{if } a_j \le c(m)y^m, \\ \binom{2\frac{a_j}{y}}{c(m)y^{m-1}} & \text{if } a_j > c(m)y^m. \end{cases}$$

Case (i): Using the upper bound $\binom{i}{r} \le \left(\frac{ie}{r}\right)^r$, for $r,i \in \mathbb{N}$ with $r \le i$ and the inequality $\left(\frac{a_j}{c(m)}\right)^{1/m} \le y \le a_j \le n$ we obtain, for some constants $C_0 = C_0(m)$, $C_1 = C_1(m)$,

$$\binom{2c(m)y^{m-1}}{a_j/y} \le \left(\frac{2c(m)y^{m-1}e}{a_j/y}\right)^{a_j/y} \le \exp\left(C_0\,(a_j/y)\log n\right) \le \exp\left((\log n)\,C_1\,a_j^{\frac{m-1}{m}}\right).$$

Case (ii): The same upper bound for binomial coefficients and $1 \le y \le \left(\frac{a_j}{c}\right)^{1/m} \le a_j \le n$ yield for some constant $C_2 = C_2(m)$,

$$\binom{2a_j/y}{c(m)y^{m-1}} \le \left(\frac{2(a_j/y)e}{c(m)y^{m-1}}\right)^{c(m)y^{m-1}} \le \exp\left((\log n)\,C_2\,a_j^{\frac{m-1}{m}}\right).$$

From this, we have for some $C = C(m) > 0$, the upper bound

$$\prod_{j=1}^{r} \binom{b(y_j,m)+k_j}{k_j} \le \max_{\substack{a_1,\ldots,a_r>0 \\ \sum a_j=n}} \prod_{j=1}^{r} \exp\left((\log n)C\,a_j^{\frac{m-1}{m}}\right),$$

which is estimated further (using Hölder's inequality) by

$$\exp\left((\log n)\,C\sum_{j=1}^{r} a_j^{\frac{m-1}{m}}\right) \le \exp\left((\log n)\,C\,r^{\frac{1}{m}}\left(\sum_{j=1}^{r} a_j\right)^{\frac{m-1}{m}}\right).$$

We observe that all y_j are different, positive and that their sum is not greater than n, so we have that

$$r^2/2 \le 1 + \ldots + r \le y_1 + \ldots + y_r \le n.$$

Recalling that $a_1 + \cdots + a_r = n$, our upper bound becomes

$$\exp\left((\log n)\,C\,(2n)^{\frac{1}{2m}}\,n^{\frac{m-1}{m}}\right) = \exp\left(\tfrac{\vartheta}{2}\,(\log n)\,n^{\frac{2m-1}{2m}}\right),$$

for some $\vartheta = \vartheta(m) > 0$. Note that from our argument so far one can easily recover the (well-known) fact that the number of integer partitions of n is bounded by $e^{(\vartheta/2)\sqrt{n}}$. Combining this with the upper bound for the number of integer partitions with a given sum-partition, we obtain the claim. ∎

Proof of Lemma 4.1   Suppose $\omega_n \in \mathcal{M}_n(\mathcal{X})$ and $\varpi_n \in \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$. For $a \in \mathcal{X}$, we look at the mappings

$$\Phi_a \colon \mathcal{K}^{(n)}(\omega_n, \varpi_n) \ni \frac{1}{n} \sum_{v \in V} \delta_{(X(v),L(v))} \mapsto \left(L_1^a, \ldots, L_{n\omega(a)}^a \right),$$

where $(L_1^a, \ldots, L_{n\omega(a)}^a)$ is the ordering of the vectors $L(v)$, for all $v \in V$ with $X(v) = a$, and thus constitutes and integer partition of the vector

$$\left(n\varpi_n(a,b) \colon b \in \mathcal{X} \right),$$

which has magnitude $n \sum_b \varpi_n(a,b)$. The combined mapping $\Phi = (\Phi_a \colon a \in \mathcal{X})$ is injective, and therefore, by Lemma 4.2,

$$\sharp \mathcal{K}^{(n)}(\omega_n, \varpi_n) \leq \exp \left[\vartheta \sum_{a \in \mathcal{X}} \log \left(n \sum_{b \in \mathcal{X}} \varpi_n(a,b) \right) \left(n \sum_{b \in \mathcal{X}} \varpi_n(a,b) \right)^{\frac{2m-1}{2m}} \right]$$
$$\leq \exp \left[\vartheta \log \left(n \|\varpi\| \right) m^{\frac{1}{2m}} \left(n \|\varpi_n\| \right)^{\frac{2m-1}{2m}} \right],$$

where we have used the fact that $\varpi_n(a,b) \leq \|\varpi\|$ and Hölder's inequality in the last step. ∎

4.2 Proof of the upper bound in Theorem 2.5.

We are now ready to prove an upper bound for the large deviation probability in Theorem 2.5, starting with a lemma based on the method of types.

Lemma 4.3. *For any sequence (ν_n) with $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ we have*

$$\mathbb{P}\{M = \nu_n \,|\, \Phi(M) = (\omega_n, \varpi_n)\} \leq \exp\left(-nH(\nu_n \,\|\, Q_n) + \varepsilon_1^{(n)}(\nu_n) \right),$$

where

$$Q_n(a,\ell) = \omega_n(a) \prod_{b \in \mathcal{X}} \frac{e^{-\varpi_n(a,b)/\omega_n(a)} [\varpi_n(a,b)/\omega_n(a)]^{\ell(b)}}{\ell(b)!}, \quad \text{for } \ell \in \mathcal{N}(\mathcal{X}),$$

and

$$\lim_{n \uparrow \infty} \frac{1}{n} \sup_{\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)} \varepsilon_1^{(n)}(\nu_n) = 0.$$

Proof.   The proof of this lemma is based on the method of types, see [DZ98, Chapter 2]. For any $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ we have

$$\mathbb{P}\{M = \nu_n \,|\, \Phi(M) = (\omega_n, \varpi_n)\} = \frac{\sharp\{\tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n) \,:\, M \circ \Psi(\tilde{x}) = \nu_n\}}{\sharp\{\tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n))\}}. \tag{4.2}$$

Now, by elementary counting, the denominator on the right side of (4.2) is

$$\binom{n}{n\omega_n(a),\, a \in \mathcal{X}} \prod_{\{a,b\}} \prod_{k=1}^{n(a,b)} \left(\frac{n^2 \omega_n(a)\omega_n(b) - n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1 + \mathbb{1}_{\{a=b\}}} - (k-1) \right). \tag{4.3}$$

For a given empirical neighbourhood measure $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ the numerator is probably too tricky to find explicitly. However, an easy upper bound is

$$\binom{n}{n\nu_n(a, \ell), \, a \in \mathcal{X}, \ell \in \mathcal{N}(\mathcal{X})} \prod_{(a,b)} \binom{n\varpi_n(a, b)}{\ell_a^{(j)}(b), \, j = 1, \ldots, n\omega_n(a)} 2^{-\frac{n}{2}\varpi_n(\Delta)}, \qquad (4.4)$$

where $\ell_a^{(j)}(b), \, j = 1, \ldots, n\omega_n(a)$ are any enumeration of the family containing each $\ell(b)$ with multiplicity $n\nu_n(a, \ell)$. This upper bound is obtained by attaching edge-numbers without discounting for the possibility of multiple edges or loops. In the case $a = b$ initially edges are considered to be oriented and then the orientation is forgotten, leading to the extra term $2^{-\frac{n}{2}\varpi_n(\Delta)}$. Combining (4.2), (4.3), and (4.4) we get

$$\mathbb{P}\{M = \nu_n \mid \Phi(M) = (\omega_n, \varpi_n)\}$$
$$\leq \prod_{a \in \mathcal{X}} \binom{n\omega_n(a)}{n\nu_n(a, \ell), \, \ell \in \mathcal{N}(\mathcal{X})} \prod_{(a,b)} \binom{n\varpi_n(a, b)}{\ell_a^{(j)}(b), \, j = 1, \ldots, n\omega_n(a)} \qquad (4.5)$$
$$\times 2^{-\frac{n}{2}\varpi_n(\Delta)} \prod_{\{a,b\}} \prod_{k=1}^{n(a,b)} \left(\tfrac{n^2\omega_n(a)\omega_n(b) - n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}} - (k-1)\right)^{-1},$$

It remains to analyse the asymptotics of the upper bound. Using Stirling's formula, we obtain

$$\prod_{a \in \mathcal{X}} \binom{n\omega_n(a)}{n\nu_n(a, \ell), \, \ell \in \mathcal{N}(\mathcal{X})} \leq \exp\left(n \sum_a \omega_n(a) \log \omega_n(a) - n \sum_{(a,\ell)} \nu_n(a, \ell) \log \nu_n(a, \ell)\right)$$
$$\times \exp\left(\tfrac{m}{2} \log(2\pi n) + \tfrac{1}{n} \sum_a \tfrac{1}{12\omega_n(a)}\right).$$

We observe that

$$\prod_{j=1}^{n\omega_n(a)} (\ell_a^{(j)}(b))! = \exp\left(n \sum_\ell \log\left(\ell(b)!\right) \nu_n(a, \ell)\right),$$

and hence

$$\binom{n\varpi_n(a, b)}{\ell_a^{(j)}(b), \, j \leq n\omega_n(a)} \leq \exp\left(-n \sum_\ell \log\left(\ell(b)!\right) \nu_n(a, \ell) + n\varpi_n(a, b) \log\left(n\varpi_n(a, b)\right) - n\varpi_n(a, b)\right)$$
$$\times \exp\left(\tfrac{1}{12n\varpi_n(a,b)} + \tfrac{1}{2} \log(2\pi n)\right).$$

Next, we obtain,

$$\prod_{k=1}^{n(a,b)} \left(\tfrac{n^2\omega_n(a)\omega_n(b) - n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}}\right) - (k-1)) \leq \exp\left(n(a, b) \log\left(\tfrac{n^2\omega_n(a)\omega_n(b)}{1+\mathbb{1}_{\{a=b\}}}\right\}\right)$$
$$\times \exp\left(n(a, b) \log\left(1 - \tfrac{\mathbb{1}_{\{a=b\}}}{2n\omega_n(a)} - \tfrac{n(a,b)}{n^2\omega_n(a)\omega_n(b)}\right)\right).$$

Putting everything together, recalling that $H(\omega)$ is the entropy of a discrete measure ω, we get

$$\mathbb{P}\{M = \nu_n \,|\, \Phi(M) = (\omega_n, \varpi_n)\}$$

$$\leq \exp\Big(-nH(\omega_n) + nH(\nu_n) - n\sum_{(a,b)}\sum_{\ell}(\log \ell(b)!)\,\nu_n(a,\ell) + n\sum_{(a,b)}\varpi_n(a,b)\log\varpi_n(a,b)$$

$$- n\sum_{(a,b)}\varpi_n(a,b) - \tfrac{n}{2}\sum_{(a,b)}\varpi_n(a,b)\,\log\big((1 + \mathbb{1}_{\{a=b\}})\,\omega_n(a)\omega_n(b)\big) + \varepsilon_1^{(n)}\Big),$$

for a sequence $\varepsilon_1^{(n)}$ which does not depend on ν_n and satisfies $\lim_{n\uparrow\infty}\frac{1}{n}\varepsilon_1^{(n)} = 0$. To give the right hand side the form as stated in the theorem, we observe that

$$H(\omega_n) - H(\nu_n) + \sum_{(a,b)}\sum_{\ell}(\log \ell(b)!)\,\nu_n(a,\ell) - \sum_{(a,b)}\varpi_n(a,b)\log\varpi_n(a,b) + \sum_{(a,b)}\varpi_n(a,b)$$

$$+ \tfrac{1}{2}\sum_{(a,b)}\varpi_n(a,b)\,\log\big((1 + \mathbb{1}_{\{a=b\}})\,\omega_n(a)\omega_n(b)\big)$$

$$= \sum_{(a,\ell)}\nu_n(a,\ell)\Big[\log\nu_n(a,\ell) - \log\omega_n(a) - \sum_b\Big(\log\big(\tfrac{\varpi_n(a,b)}{\omega_n(a)}\big)^{\ell(b)} - \tfrac{\varpi_n(a,b)}{\omega_n(a)} - (\log \ell(b)!)\Big)\Big]$$

$$= \sum_{(a,\ell)}\nu_n(a,\ell)\Big[\log\nu_n(a,\ell) - \log\big(\omega_n(a)\prod_b \tfrac{(\varpi_n(a,b)/\omega_n(a))^{\ell(b)}\exp(-\varpi_n(a,b)/\omega_n(a))}{\ell(b)!}\big)\Big]$$

$$= H(\nu_n \,\|\, Q_n),$$

which completes the proof of Lemma 4.3. ∎

We can now complete the proof of the upper bound in Theorem 2.5 by combining Lemma 4.1 and Lemma 4.3. Suppose that $\Gamma \subset \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ is a closed set. Then,

$$\mathbb{P}\Big\{M \in \Gamma \,\Big|\, \Phi(M) = (\omega_n, \varpi_n)\Big\} = \sum_{\nu_n \in \Gamma\cap\mathcal{K}^{(n)}(\omega_n,\varpi_n)}\mathbb{P}\big\{M = \nu_n \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\}$$

$$\leq \sharp\mathcal{K}^{(n)}(\omega_n,\varpi_n)\exp\Big(-n\inf_{\nu_n \in \Gamma\cap\mathcal{K}^{(n)}(\omega_n,\varpi_n)}H(\nu_n \,\|\, Q_n) + \sup_{\nu_n \in \mathcal{K}^{(n)}(\omega_n,\varpi_n)}\varepsilon_1^{(n)}(\nu_n)\Big).$$

We have already seen that $\frac{1}{n}\sup_{\nu_n}\varepsilon_1^{(n)}(\nu_n)$ and $\frac{1}{n}\log\sharp\mathcal{K}^{(n)}(\omega_n,\varpi_n)$ converge to zero. It remains to check that

$$\lim_{n\to\infty}\sup_{\nu_n \in \mathcal{K}^{(n)}(\omega_n,\varpi_n)}\big|H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q)\big| = 0. \tag{4.6}$$

To do this, we observe that

$$H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q) = \sum_{(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})}\nu_n(a,\ell)\log\tfrac{Q(a,\ell)}{Q_n(a,\ell)}$$

$$= -H(\omega_n \,\|\, \omega) - H(\varpi_n \,\|\, \varpi) - \sum_{a,b\in\mathcal{X}}\varpi(a,b)\tfrac{\omega_n(a)}{\omega(a)} + \sum_{a,b\in\mathcal{X}}\varpi(a,b)\log\tfrac{\omega_n(a)}{\omega(a)} + \|\varpi_n\|. \tag{4.7}$$

Note that this expression does not depend on ν_n. As the first, second and fourth term of (4.7) converge to 0, and the third and fifth term converge to $\|\varpi\|$, the expression (4.7) vanishes in the limit, and this completes the proof of the upper bound in Theorem 2.5.

4.3 An upper bound on the support of empirical neighbourhood measures

The support, denoted $\mathcal{S}(\nu) \subset \mathcal{X} \times \mathcal{N}(\mathcal{X})$, of an empirical neighbourhood measure ν of a graph with n vertices is naturally bounded by n. For the proof of the lower bound in Theorem 2.5 we need a better upper bound. We still use m to denote the cardinality of $\mathcal{X}$, and let

$$C := 2^m \frac{\Gamma(m+2)^{\frac{m}{m+1}}}{\Gamma(m)} \text{ and } D := 2^m \frac{(m+1)^m}{\Gamma(m)},$$

where $\Gamma(\cdot)$ is the Gamma function.

Lemma 4.4. *For every* $(\omega_n, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$ *and* $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ *with* $\Phi(\nu_n) = (\omega_n, \varpi_n)$, *we have*

$$\sharp \mathcal{S}(\nu_n) \leq C \left[n\|\varpi_n\| \right]^{\frac{m}{m+1}} + D. \tag{4.8}$$

The following lemma provides a step in the proof of Lemma 4.4.

Lemma 4.5. *Suppose* $j \in \mathbb{N} \cup \{0\}$ *and* $n \in \mathbb{N}$. *Then,*

$$\tfrac{1}{\Gamma(n)} j^{n-1} \leq \sharp \Big\{ (l_1, \ldots, l_n) \in \big(\mathbb{N} \cup \{0\}\big)^n : l_1 + \ldots + l_n = j \Big\} \leq \tfrac{1}{\Gamma(n)} (j+n)^{n-1}. \tag{4.9}$$

Proof.   The proof is by induction on n. Equation (4.9) holds trivially for all $j \in \mathbb{N} \cup \{0\}$ and $n = 1, 2$, so we assume it holds for all j and $n \geq 2$. By the induction hypothesis, for any j,

$$\frac{1}{\Gamma(n)} \sum_{l=0}^{j} (j-l)^{n-1} \leq \sum_{l=0}^{j} \sharp \Big\{ (l_1, \ldots, l_{n-1}) \in \big(\mathbb{N} \cup \{0\}\big)^{n-1} : l_1 + \ldots + l_{n-1} = j - l \Big\}$$

$$= \sharp \Big\{ (l_1, \ldots, l_n) \in \big(\mathbb{N} \cup \{0\}\big)^n : l_1 + \ldots + l_n = j \Big\} \leq \frac{1}{\Gamma(n)} \sum_{l=0}^{j} (j-l+n)^{n-1}.$$

For the first and last term, we obtain the lower and upper bounds

$$\sum_{l=0}^{j} (j-l)^{n-1} \geq \int_0^j y^{n-1}\, dy = \frac{1}{n} j^n = \frac{\Gamma(n)}{\Gamma(n+1)} j^n$$

and

$$\sum_{l=0}^{j} (j-l+n)^{n-1} \leq \int_n^{j+n} y^{n-1} dy \leq \int_0^{j+n+1} y^{n-1} dy = \frac{1}{n}(j+n+1)^n = \frac{\Gamma(n)}{\Gamma(n+1)}(j+n+1)^n,$$

which yields inequality (4.9) for $n+1$ instead of n, and completes the induction. ∎

Proof of Lemma 4.4.   Suppose $(\omega_n, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \mathcal{M}_n(\mathcal{X} \times \mathcal{X})$. Let

$$a_m(j) := \sharp \Big\{ (a, \ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X}) : \sum_{b \in \mathcal{X}} \ell(b) = j \Big\}$$

$$= m \times \sharp \Big\{ (l_1, \ldots, l_m) \in \big(\mathbb{N} \cup \{0\}\big)^m : l_1 + \ldots + l_m = j \Big\}.$$

For any positive integer k we write

$$\theta_k = \min \Big\{ \theta \in \mathbb{N} : \sum_{j=0}^{\theta} a_m(j) \geq k \Big\}.$$

We observe from Lemma 4.5 that,

$$k \leq \sum_{j=0}^{\theta_k} a_m(j) \leq m \sum_{j=0}^{\theta_k} \tfrac{1}{\Gamma(m)}(j+m)^{m-1} \leq \tfrac{m}{\Gamma(m)} \int_0^{\theta_k+m} y^{m-1} dy = \tfrac{1}{\Gamma(m)}\big(\theta_k + m\big)^m.$$

Thus, we have $\theta_k \geq \big(k\Gamma(m)\big)^{\frac{1}{m}} - m =: \alpha_k$. This yields

$$\sum_{j=0}^{\theta_k} j a_m(j) \geq \tfrac{1}{\Gamma(m)} \sum_{j=0}^{\lceil \alpha_k \rceil} j^m \geq \tfrac{1}{\Gamma(m)} \int_0^{\alpha_k - 1} y^m dy \geq \tfrac{1}{\Gamma(m+2)}\big(\alpha_k - 1\big)^{m+1}, \qquad (4.10)$$

where $\lceil y \rceil$ is the smallest integer greater or equal to y.

Observe that the size of the support of the measure $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ satisfies

$$\sharp \mathcal{S}(\nu_n) \leq \max \Big\{ k \colon \sum_{j=0}^{\theta_k} j a_m(j) \leq n\|\varpi_n\| \Big\},$$

and hence, using (4.10) and the inequality $(a+b)^m \leq 2^m(a^m + b^m)$ for $a, b \geq 0$,

$$\sharp \mathcal{S}(\nu_n) \leq \max \Big\{ k \colon \tfrac{1}{\Gamma(m+2)}\big(\alpha_k - 1\big)^{m+1} \leq n\|\varpi_n\| \Big\}$$

$$\leq \Gamma(m)^{-1} \Big((n\|\varpi_n\|)^{\frac{1}{m+1}} \Gamma(m+2)^{\frac{1}{m+1}} + m + 1 \Big)^m \leq C\, (n\|\varpi_n\|)^{\frac{m}{m+1}} + D,$$

where the constants C, D were define before the formulation of the lemma. ∎

4.4 Approximation by empirical neighbourhood measures

Throughout this section we assume that $\omega_n \in \mathcal{M}_n(\mathcal{X})$ with $\omega_n \to \omega$, $\varpi_n \in \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$ with $\varpi_n \to \varpi$, and that $\nu \in \mathfrak{M}[\omega, \varpi]$. Our aim is to show that ν can be approximated in the weak topology by the empirical neighbourhood measure of a graph with n vertices, empirical colour measure ω_n, empirical pair measure ν_n, and the additional feature that the degree of any vertex is bounded by $n^{1/3}$. The approximation will be done in four steps.

We denote by d the metric of total variation, i.e.

$$d(\nu, \tilde{\nu}) = \tfrac{1}{2} \sum_{(a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})} |\nu(a, \ell) - \tilde{\nu}(a, \ell)|, \quad \text{for } \nu, \tilde{\nu} \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})).$$

This metric generates the weak topology.

Lemma 4.6 (Approximation Step 1). *For every $\varepsilon > 0$, there exist $\hat{\nu} \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ and $\hat{\varpi} \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ such that $|\varpi(a, b) - \hat{\varpi}(a, b)| \leq \varepsilon$ for all $a, b \in \mathcal{X}$, $d(\nu, \hat{\nu}) \leq \varepsilon$ and $(\hat{\varpi}, \hat{\nu})$ is consistent.*

Proof. By our assumption (ϖ, ν) is sub-consistent. For any $b \in \mathcal{X}$ define $e^{(b)} \in \mathcal{N}(\mathcal{X})$ by $e^{(b)}(a) = 0$ if $a \neq b$, and $e^{(b)}(b) = 1$. For large n define measures $\hat{\nu}_n \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ by

$$\hat{\nu}_n(a, \ell) = \nu(a, \ell)\Big(1 - \tfrac{\|\varpi\| - \|\langle \nu(\cdot, \ell), \ell(\cdot)\rangle\|}{n}\Big) + \sum_{b \in \mathcal{X}} \mathbb{1}\{\ell = ne^{(b)}\}\, \tfrac{\varpi(a,b) - \langle \nu(\cdot, \ell), \ell(\cdot)\rangle(a,b)}{n}.$$

Note that $\hat{\nu}_n \to \nu$ and that, for all $a, b \in \mathcal{X}$,

$$\sum_{\ell \in \mathcal{N}(\mathcal{X})} \hat{\nu}_n(a, \ell)\ell(b) = \Big(1 - \tfrac{\|\varpi\| - \|\langle \nu(\cdot, \ell), \ell(\cdot)\rangle\|}{n}\Big) \sum_{\ell \in \mathcal{N}(\mathcal{X})} \nu(a, \ell)\ell(b) + \varpi(a, b) - \langle \nu(\cdot, \ell), \ell(\cdot)\rangle(a, b)$$

$$= \varpi(a, b) - \tfrac{\|\varpi\| - \|\langle \nu(\cdot, \ell), \ell(\cdot)\rangle\|}{n} \langle \nu(\cdot, \ell), \ell(\cdot)\rangle(a, b) \xrightarrow{n\uparrow\infty} \varpi(a, b).$$

Hence, defining $\hat{\varpi}_n$ by $\hat{\varpi}_n(a,b) = \sum \hat{\nu}_n(a,\ell)\ell(b)$, we have a sequence of consistent pairs $(\hat{\varpi}_n, \hat{\nu}_n)$ converging to (ϖ, ν), as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 4.7 (Approximation Step 2). *For every $\varepsilon > 0$, there exists $n(\varepsilon)$ such that, for all $n \geq n(\varepsilon)$ there exists $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n, \varpi_n)$ such that $d(\nu_n, \nu) \leq \varepsilon$.*

The key to the construction of the measure ν_n is the following 'law of large numbers'.

Lemma 4.8. *For every $\delta > 0$, there exists $\hat{\nu} \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $d(\nu, \hat{\nu}) < \delta$ such that, for i.i.d. $\mathcal{N}(\mathcal{X})$-valued random variables ℓ_j^a, $j = 1, \ldots, n\omega_n(a)$ with law $\hat{\nu}(\,\cdot\,|\,a) := \hat{\nu}(\{a\} \times \cdot)/\hat{\nu}_1(a)$, almost surely,*

$$\limsup_{n\to\infty} \Big(\frac{1}{n} \sum_{j=1}^{n\omega_n(a)} \ell_j^a(b) - \varpi_n(a,b) \Big) \leq \delta, \qquad \text{for all } a,b \in \mathcal{X}. \qquad (4.11)$$

Proof.   By Lemma 4.6 we can choose a consistent pair $(\hat{\varpi}, \hat{\nu})$ such that $d(\nu, \hat{\nu}) < \delta$ and, for all $a, b \in \mathcal{X}$,

$$\frac{\nu_1(a)}{\hat{\nu}_1(a)} \leq 1 + \frac{\delta}{\|\varpi\|+1} \quad \text{and} \quad \hat{\varpi}(a,b)\big(1 + \frac{\delta}{\|\varpi\|+1}\big) \leq \varpi(a,b)\big(1 + \frac{\delta}{\varpi(a,b)}\big).$$

The random variables $\ell_j^a(b)$, $j = 1, \ldots, n\omega_n(a)$ are i.i.d. with expectation

$$\mathbb{E}\ell_1^a(b) = \sum_\ell \hat{\nu}(\ell\,|\,a)\ell(b) = \frac{\hat{\varpi}(a,b)}{\hat{\nu}_1(a)}.$$

Hence, by the strong law of large numbers, almost surely,

$$\limsup_{n\to\infty} \Big(\frac{1}{n} \sum_{j=1}^{n\omega_n(a)} \ell_j^a(b) - \varpi_n(a,b) \Big) \leq \frac{\nu_1(a)}{\hat{\nu}_1(a)}\hat{\varpi}(a,b) - \varpi(a,b) \leq \delta,$$

where we also used that $\omega_n(a) \to \omega(a) = \nu_1(a)$ and $\varpi_n(a,b) \to \varpi(a,b)$. $\qquad\qquad\blacksquare$

Proof of Lemma 4.7.   We use a random construction. Given $\nu \in \mathfrak{M}[\omega, \varpi]$ and $\varepsilon > 0$, choose $\hat{\nu}$ as in Lemma 4.8 with $\delta = \varepsilon/(3m)$, where m is the cardinality of $\mathcal{X}$. For every $a \in \mathcal{X}$, we draw tuples ℓ_j^a, $j = 1, \ldots, n\omega_n(a)$ independently according to $\hat{\nu}(\,\cdot\,|\,a)$ and define $e_n(a,b)$ by

$$e_n(a,b) := \frac{1}{n} \sum_{j=1}^{n\omega_n(a)} \ell_j^a(b) - \varpi_n(a,b), \ \text{ for all } a,b \in \mathcal{X}.$$

We modify the tuples $(\ell_j^a : j = 1, \ldots, n\omega_n(a))$ as follows:

- If $e_n(a,b) < 0$, we add an amount to the last element $\ell_{n\omega_n(a)}^a(b)$ such that the modified tuple satisfies $e_n(a,b) = 0$;
- if $e_n(a,b) > 0$, by Lemma 4.8, the 'overshoot' $ne_n(a,b)$ cannot exceed $n\delta$. We successively deduct one from the nonzero elements in $\ell_j^a(b)$, $j = 1, \ldots, n\omega_n(a)$ until the modified tuples satisfy $e_n(a,b) = 0$;
- if $e_n(a,b) = 0$ we do not modify $\ell_j^a(b)$.

We denote by $(\tilde{\ell}_j^a : j = 1, \ldots, n\omega_n(a))$ the tuples after all modifications.

For each $a \in \mathcal{X}$ define probability measures $\tilde{\Delta}_n(\,\cdot\,|\,a)$ and $\Delta_n(\,\cdot\,|\,a)$ by

$$\tilde{\Delta}_n(\ell\,|\,a) = \frac{1}{n\omega_n(a)} \sum_{j=1}^{n\omega_n(a)} \mathbb{1}_{\{\tilde{\ell}_j^a = \ell\}}, \qquad \text{for } \ell \in \mathcal{N}(\mathcal{X}),$$

respectively,

$$\Delta_n(\ell\,|\,a) = \frac{1}{n\omega_n(a)} \sum_{j=1}^{n\omega_n(a)} \mathbb{1}_{\{\ell_j^a = \ell\}}, \qquad \text{for } \ell \in \mathcal{N}(\mathcal{X}).$$

We define probability measures $\tilde{\nu}_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ and $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ by $\tilde{\nu}_n(a, \ell) = \omega_n(a)\Delta_n(\ell\,|\,a)$, respectively $\nu_n(a, \ell) = \omega_n(a)\tilde{\Delta}_n(\ell\,|\,a)$, for $(a, \ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})$. Recall from our modification procedure that, in the worst case, we have changed $nm\delta$ of the tuples. Thus,

$$d(\tilde{\nu}_n, \nu_n) \le m\delta \le \tfrac{1}{3}\varepsilon\,.$$

As a result of our modifications we have $\Phi(\nu_n) = (\omega_n, \varpi_n)$. We observe that, for all $(a, \ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})$, the random variables

$$\mathbb{1}\{\ell_1^a = \ell\}, \ldots, \mathbb{1}\{\ell_{n\omega_n(a)}^a = \ell\}$$

are independent Bernoulli random variables with success probability $\hat{\nu}(\ell\,|\,a)$ and hence, almost surely,

$$\lim_{n\to\infty} \Delta_n(\ell\,|\,a) = \hat{\nu}(\ell\,|\,a)\,.$$

Therefore, for all $(a, \ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})$, we obtain $\lim_{n\to\infty} \tilde{\nu}_n(a, \ell) = \hat{\nu}(a, \ell)$, almost surely. Thus, almost surely, for all large n, we have $d(\nu_n, \nu) \le d(\nu_n, \tilde{\nu}_n) + d(\tilde{\nu}_n, \hat{\nu}) + d(\hat{\nu}, \nu) \le \varepsilon$, as claimed. ∎

Lemma 4.9 (Approximation Step 3). *Let* $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ *with* $\Phi(\nu_n) = (\omega_n, \varpi_n)$. *For every* $\varepsilon > 0$ *there exists* $n(\varepsilon)$ *such that, for all* $n \ge n(\varepsilon)$, *we can find* $\tilde{\nu}_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ *with* $\Phi(\tilde{\nu}_n) = (\omega_n, \varpi_n)$ *and* $d(\nu_n, \tilde{\nu}_n) < \varepsilon$, *such that*

$$\sum_{b \in \mathcal{X}} \ell(b) \le n^{1/3} \text{ for } \tilde{\nu}_n\text{-almost every } (a, \ell). \tag{4.12}$$

Proof. As $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$, there is a representation

$$\nu_n = \frac{1}{n} \sum_{k=1}^n \delta_{(a_k, \ell_k)}, \qquad \text{for } a_k \in \mathcal{X}, \ell_k \in \mathcal{N}(\mathcal{X}).$$

Fix $\delta > 0$ and $a \in \mathcal{X}$. Look at the sets

- $V^+ = \{1 \le k \le n : a_k = a, \sum_b \ell_k(b) > n^{1/3}\}$ with $\sharp V^+ \le (n\sum_b \varpi_n(a, b))^{2/3}$,
- $V^- = \{1 \le k \le n : a_k = a, \sum_b \ell_k(b) \le n^{1/4}\}$ with $\sharp V^- \ge n - (n\sum_b \varpi_n(a, b))^{3/4}$.

For each $k \in V_1$ we replace ℓ_k by a smaller vector $\tilde{\ell}_k$ such that $\sum_b \tilde{\ell}_k(b) = n^{1/3}$. As

$$\sum_{k \in V_1} \sum_b \ell_k(b) \le n \sum_{b \in \mathcal{X}} \varpi_n(a, b)$$

we replace (for large n) no more than δn of the vectors ℓ_k, $k \in V_2$, by larger vectors $\tilde{\ell}_k$ such that

$$\sum_b \tilde{\ell}_k(b) \le n^{1/3} \text{ and } \sum_{k=1}^n \sum_{b \in \mathcal{X}} \mathbb{1}\{a_k = a\}\,\tilde{\ell}_k(b) = \sum_{k=1}^n \sum_{b \in \mathcal{X}} \mathbb{1}\{a_k = a\}\,\ell_k(b)\,,$$

where we use the convention $\tilde{\ell}_k = \ell_k$ if this vector was not changed in the procedure. Performing such an operation for every $a \in \mathcal{X}$ we may define

$$\tilde{\nu}_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{(a_k, \tilde{\ell}_k)},$$

and observe that (4.12) holds and $\Phi(\tilde{\nu}_n) = (\omega_n, \varpi_n)$. Moreover,

$$d(\nu_n, \tilde{\nu}_n) \leq \tfrac{m}{2n}\big((n \sum_b \varpi_n(a,b))^{2/3} + \delta n\big),$$

which is less than $\varepsilon > 0$ for a suitable choice of $\delta > 0$, and all sufficiently large n. ∎

Lemma 4.10 (Approximation Step 4). *Let $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n, \varpi_n)$. For every $\varepsilon > 0$ there exists $n(\varepsilon)$ such that, for all $n \geq n(\varepsilon)$, we can find $\tilde{\nu}_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ with $d(\nu_n, \tilde{\nu}_n) < \varepsilon$ such that (4.12) holds.*

Proof.   By Lemma 4.9 we may assume that ν_n can be represented as

$$\nu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{(a_k, \ell_k)}, \qquad \text{for } a_k \in \mathcal{X}, \ell_k \in \mathcal{N}(\mathcal{X}) \text{ with } \sum_{b \in \mathcal{X}} \ell_k(b) \leq n^{1/3}.$$

For the proof it suffices to construct, for every $\varepsilon > 0$ and all large n, a (coloured) random graph X with n vertices such that

$$\limsup_{n \to \infty} \mathbb{P}\big\{d(\nu_n, M(X)) \geq \varepsilon\big\} = 0, \tag{4.13}$$

We now describe the random procedure that generates X. First, equip each vertex with an element of $\mathcal{X} \times \mathcal{N}(\mathcal{X})$ by drawing without replacement from the collection $\{(a_1, \ell_1), \ldots, (a_n, \ell_n)\}$. We denote by $V(a)$ the collection of vertices which have colour $a \in \mathcal{X}$, an observe that $\sharp V(a) = n\omega_n(a)$.

Now fix $a, b \in \mathcal{X}$. We construct two collections of vertices: If a vertex is equipped with (a_k, ℓ_k) and $a_k = a$, then it is represented $\ell_k(b)$ times in the first collection, $W(a)$. If $a_k = b$, then it is represented $\ell_k(b)$ times in the second collection $W(b)$. Hence there are exactly $n\varpi_n(a,b)/2$ vertices of colour a in collection $W(a)$, and exactly $n\varpi_n(a,b)/2$ vertices of colour b in collection $W(b)$.

We now match vertices from the two collections randomly: At each step $k = 1, \ldots, n\varpi_n(a,b)/2$, we randomly pick two vertices $V_1^k \in W(a)$ and $V_2^k \in W(b)$. We connect V_1^k and V_2^k by an edge unless $V_1^k = V_2^k$ or the two vertices are already connected. If one of these two things happen, then we simply choose an edge randomly from the set of all possible edges connecting colours a and b, which are not yet present in the graph and whose introduction does not violate (4.12).

This completes the construction of a graph with $L^1(X) = \omega_n$, $L^2(X) = \varpi_n$ and

$$d(\nu_n, M(X)) \leq \tfrac{2}{n} \sum_{a,b \in \mathcal{X}} B^n(a,b), \tag{4.14}$$

where $B^n(a,b)$ is the total number of steps $k \in \{1, \ldots, n\varpi_n(a,b)/2\}$ at which there is disparity between the vertices V_1^k, V_2^k drawn and the vertices which formed the kth edge connecting a and b in the random graph construction.

Given $a, b \in \mathcal{X}$, using $\sum \ell_k(b) \leq n^{1/3}$, the probability that $V_1^k = V_2^k$ or the two vertices are already connected is less or equal

$$p_{[k]}(a,b) := \frac{2n^{1/3}}{n\varpi_n(a,b)} \mathbb{1}_{\{a=b\}} + \big(1 - \frac{2n^{1/3}}{n\varpi_n(a,b)} \mathbb{1}_{\{a=b\}}\big) \frac{4(k-1)n^{2/3}}{(n\varpi_n(a,b))^2}.$$

$B^n(a,b)$ is a sum of independent Bernoulli random variables $X_1, \ldots, X_{n\varpi_n(a,b)/2}$ with 'success' probabilities less or equal $p_{[1]}(a,b), \ldots, p_{[n\varpi_n(a,b)/2]}(a,b)$. So its expectation satisfies

$$\mathbb{E}B^n(a,b) \leq \sum_{k=1}^{n\varpi_n(a,b)/2} k p_{[k]}(a,b) = n^{1/3}\mathbb{1}_{\{a=b\}} + n^{2/3}\big(1 - \tfrac{2}{n^{2/3}\varpi_n(a,b)}\big)\big(1 - \tfrac{2}{n\varpi_n(a,b)}\big).$$

Hence, we have that $\mathrm{Var}(B^n(a,b)) \leq \mathbb{E}B^n(a,b) = o(n)$. By Bernstein's inequality, see for example [BBL04, Theorem 3], we obtain, for any $\epsilon > 0$,

$$\mathbb{P}\big\{B^n(a,b) \geq \mathbb{E}B^n(a,b) + 2n\epsilon\big\} \leq \exp\Big\{ - \tfrac{n^2\epsilon^2}{2\,(\mathrm{Var}(B^n(a,b))+n\epsilon/3)} \Big\}.$$

Using the bounds on $\mathbb{E}B^n(a,b)$ and $\mathrm{Var}(B^n(a,b))$, we obtain, for sufficiently large n,

$$\mathbb{P}\big\{B^n(a,b) \geq o(n) + 2n\epsilon\big\} \leq \exp\Big\{ - \tfrac{n^2\epsilon^2}{2o(n)+2n\epsilon/3} \Big\}.$$

Now suppose a small $\varepsilon > 0$ and large $A > 0$ are given. Let $\delta = \varepsilon/(2m^2)$. Suppose that $B^n(a,b) \leq n\delta$, for all $a, b \in \mathcal{X}$. Then, by (4.14), $d(M(X), \nu_n) \leq 2\delta m^2 = \varepsilon$. Hence,

$$\mathbb{P}\big\{d(M(X), \nu_n) > \varepsilon\big\} \leq \sum_{a,b \in \mathcal{X}} \mathbb{P}\big\{B^n(a,b) \geq n\delta\big\} \leq m^2 \sup_{a,b \in \mathcal{X}} \mathbb{P}\big\{B^n(a,b) \geq o(n) + (n\delta)/2\big\}$$

$$\leq m^2 \exp\Big\{ - \tfrac{n^2\delta^2}{8o(n)+(4n\delta)/3} \Big\}.$$

This completes the proof of the lemma. ∎

4.5 Proof of the lower bound in Theorem 2.5.

There is a partial analogue to Lemma 4.3 for the lower bounds.

Lemma 4.11. *For any sequence (ν_n) with $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$, which satisfies (4.12), and $\varepsilon > 0$, we have*

$$\mathbb{P}\big\{d(M, \nu_n) < \varepsilon \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\} \geq \exp\big(- nH(\nu_n \,\|\, Q_n) - \varepsilon_2^{(n)}(\nu_n)\big),$$

where Q_n is as Lemma 4.3 and

$$\lim_{n\uparrow\infty} \tfrac{1}{n} \varepsilon_2^{(n)}(\nu_n) = 0.$$

Proof of Lemma 4.11.   We use the notation and some results from the proof of the upper bound, Lemma 4.3. Recall that

$$\mathbb{P}\big\{M = \nu_n \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\} = \frac{\sharp\big\{\tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n) \,:\, M \circ \Psi(\tilde{x}) = \nu_n\big\}}{\sharp\big\{\tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n))\big\}},$$

and that the denominator was evaluated in (4.3) as

$$\binom{n}{n\omega_n(a),\, a \in \mathcal{X}} \prod_{\{a,b\}} \prod_{k=1}^{n(a,b)} \big(\tfrac{n^2\omega_n(a)\omega_n(b)-n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}} - (k-1)\big).$$

The numerator can be estimated from below by assigning edge-numbers to the vertices in a manner cautious to avoid loops and double edges. In each step the number of assignments, which lead to multiple edges or loops, is bounded by the square of the maximal vertex degree and hence by $n^{2/3}$. Hence the numerator is bounded from below by

$$\binom{n}{n\nu_n(a,\ell),\, a \in \mathcal{X},\, \ell \in \mathcal{N}(\mathcal{X})} \prod_{(a,b)} \frac{(n\varpi_n(a,b) - n^{2/3})!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(b))!} \, 2^{-\frac{n}{2}\varpi_n(\Delta)}.$$

We again use Stirling's formula as in the proof of the upper bound. For the denominator we get the same main terms as in Lemma 4.3 with slightly different error terms, which however do not depend on ν_n. More interestingly, we have

$$\prod_{a \in \mathcal{X}} \binom{n\omega_n(a)}{n\nu_n(a,\ell), \, \ell \in \mathcal{N}(\mathcal{X})} \geq \exp\left(n \sum_a \omega_n(a) \log \omega_n(a) - n \sum_{(a,\ell)} \nu_n(a,\ell) \log \nu_n(a,\ell) \right)$$
$$\times \exp\left(- \frac{|\mathcal{S}(\nu_n)|}{2} \log(2\pi n) - \sum_{\substack{(a,\ell) \\ n\nu_n(a,\ell) \geq 1}} \frac{1}{12 n \nu_n(a,\ell)} \right),$$

where the exponent in the error term is of order $o(n)$, by the bound on the size of the support of ν_n given in Lemma 4.4. Further,

$$\frac{(n\varpi_n(a,b) - n^{2/3})!}{\prod_{j=1}^{n\omega_n(a)} (\ell_a^{(j)}(b))!} \geq \exp\left(-n \sum_\ell \log\left(\ell(b)!\right) \nu_n(a,\ell) + n\varpi_n(a,b) \log\left(n\varpi_n(a,b)\right) - n\varpi_n(a,b) \right)$$
$$\times \exp\left(-n^{2/3} \log\left(n\varpi_n(a,b)\right) + n\varpi_n(a,b) \log\left(1 - \frac{n^{2/3}}{n\varpi_n(a,b)}\right) \right),$$

and the result follows by combining this with facts discussed in the context of the upper bound. ∎

To complete the proof of the lower bound in Theorem 2.5, take an open set $\Gamma \subset \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$. Then, for any $\nu \in \Gamma \cap \mathfrak{M}[\omega, \varpi]$ we may find $\varepsilon > 0$ with the ball around ν of radius $2\varepsilon > 0$ contained in Γ. By our approximation, Lemmas 4.7 and 4.10, we may find $\nu_n \in \Gamma \cap \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ with $d(\nu_n, \nu) \downarrow 0$ such that (4.12) holds. Hence, for all large $n \geq n(\varepsilon)$,

$$\mathbb{P}\left\{ M \in \Gamma \,\middle|\, \Phi(M) = (\omega_n, \varpi_n) \right\} \geq \mathbb{P}\left\{ d(\nu_n, M) < \varepsilon \,\middle|\, \Phi(M) = (\omega_n, \varpi_n) \right\}$$
$$\geq \exp\left(-n H(\nu_n \,\|\, Q_n) + \varepsilon_2^{(n)}(\nu_n) \right).$$

We observe that

$$\lim_{n \to \infty} H(\nu_n \,\|\, Q_n) - H(\nu \,\|\, Q) = \lim_{n \to \infty} H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q) + \lim_{n \to \infty} H(\nu_n \,\|\, Q) - H(\nu \,\|\, Q) = 0,$$

where the last term vanishes by continuity of the entropy, and the first term was shown to vanish in the proof of Lemma 4.11. As $\nu \in \Gamma \cap \mathfrak{M}[\omega, \varpi]$ was arbitrary, this completes the proof of Theorem 2.5.

5. Proof of Theorem 2.1 by mixing

We denote by $\Theta_n := \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$ and $\Theta := \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$. With

$$P^{(n)}_{(\omega_n, \varpi_n)}(\nu_n) := \mathbb{P}\left\{ M = \nu_n \,\middle|\, \Phi(M) = (\omega_n, \varpi_n) \right\},$$
$$P^{(n)}(\omega_n, \varpi_n) := \mathbb{P}\left\{ (L^1, L^2) = (\omega_n, \varpi_n) \right\}$$

the joint distribution of L^1, L^2 and M is the mixture of $P^{(n)}_{(\omega_n, \varpi_n)}$ with $P^{(n)}(\omega_n, \varpi_n)$ defined as

$$d\tilde{P}^n(\omega_n, \varpi_n, \nu_n) := dP^{(n)}_{(\omega_n, \varpi_n)}(\nu_n) \, dP^{(n)}(\omega_n, \varpi_n). \tag{5.1}$$

Biggins [Bi04, Theorem 5(b)] gives criteria for the validity of large deviation principles for the mixtures and for the goodness of the rate function if individual large deviation principles are known. The following two lemmas ensure validity of these conditions.

Lemma 5.1. $(\tilde{P}^n \colon n \in \mathbb{N})$ *is exponentially tight.*

Proof.   Given $k \in \mathbb{N}$, we observe from Lemma 3.2 that there exists $N(k) \in \mathbb{N}$ such that, for all sufficiently large n,

$$\mathbb{P}\{M(\{\|\ell\| \geq 2kN(k)\}) \geq k^{-1} \text{ or } \|L^2\| \geq 2N(k)\} \leq \mathbb{P}\{|E| \geq nN(k)\} \leq e^{-kn}.$$

Now, for any $\theta > 0$, we define the set Ξ_θ by

$$\Xi_\theta := \big\{(\varpi, \nu) \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) : \nu\{\|\ell\| > 2lN(l)\} < l^{-1} \, \forall l \geq \theta \text{ and } \|\varpi\| < 2N(\theta)\big\}.$$

As $\{\|\ell\| \leq 2lN(l)\} \subset \mathcal{N}(\mathcal{X})$ is finite, hence compact, the set Ξ_θ is relatively compact in the weak topology, by Prohorov's criterion. Moreover, we have that

$$\tilde{P}^{(n)}\big((\Xi_\theta)^{\mathrm{c}}\big) \leq \mathbb{P}\{\|L^2\| \geq 2N(\theta)\} + \sum_{l=\theta}^{\infty} \mathbb{P}\big\{M(\{\|\ell\| > 2lN(l)\}) \geq l^{-1}\big\} \leq C(\theta)\, e^{-n\theta}.$$

Therefore, $\limsup\limits_{n\to\infty} \frac{1}{n} \log \tilde{P}^n((\mathrm{cl}\,\Xi_\theta)^{\mathrm{c}}) \leq -\theta$, which completes the proof, as $\theta > 0$ was arbitrary.   ∎

Define the function

$$\tilde{J} \colon \Theta \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) \to [0, \infty], \qquad \tilde{J}((\omega, \varpi), \nu) = \tilde{J}_{(\omega, \varpi)}(\nu).$$

Lemma 5.2. $\tilde{J}$ *is lower semicontinuous.*

Proof.   Suppose $\theta_n := ((\omega_n, \varpi_n), \nu_n)$ converges to $\theta := ((\omega, \varpi), \nu)$ in $\Theta \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$. There is nothing to show if $\liminf_{\theta_n \to \theta} \tilde{J}(\theta_n) = \infty$. Otherwise, if (ϖ_n, ν_n) is sub-consistent for infinitely many n, then

$$\varpi(a, b) = \lim_{n\uparrow\infty} \varpi_n(a, b) \geq \liminf_{n\uparrow\infty} \langle \nu_n(\cdot, \ell), \ell(\cdot)\rangle(a, b) \geq \langle \nu(\cdot, \ell), \ell(\cdot)\rangle(a, b),$$

hence (ϖ, ν) is sub-consistent. Similarly, if the first marginal of ν_n is ω_n, we see that the first marginal of ν is ω. We may therefore argue as in (4.6) to obtain

$$\liminf_{\theta_n \to \theta} J(\theta_n) = \liminf_{\theta_n \to \theta} H(\nu_n \,\|\, Q_n) \geq \lim_{\theta_n \to \theta} H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q) + \liminf_{\nu_n \to \nu} H(\nu_n \,\|\, Q) = H(\nu \,\|\, Q),$$

where the last step is because of continuity of the entropy. This proves the lemma.   ∎

By [Bi04, Theorem 5(b)] the two previous lemmas and the large deviation principles we have established in Theorem 2.3 and 2.5 ensure that under $(\tilde{P}^n)$ the random variables $(\omega_n, \varpi_n, \nu_n)$ satisfy a large deviation principle on $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with good rate function

$$\hat{J}(\omega, \varpi, \nu) = \begin{cases} H(\omega \,\|\, \mu) + \frac{1}{2}\,\mathfrak{H}_C(\varpi \,\|\, \omega) + H(\nu \,\|\, Q), & \text{if } \nu \in \mathfrak{M}[\omega, \varpi], \\ \infty, & \text{otherwise.} \end{cases}$$

By projection onto the last two components we obtain the large deviation principle as stated in Theorem 2.1 from the contraction principle, see e.g. [DZ98, Theorem 4.2.1].

6. Proof of Corollaries 2.2 and 2.4

We derive the corollaries from Theorem 2.1 by applying the contraction principle, see e.g. [DZ98, Theorem 4.2.1]. It just remains to simplify the rate functions.

6.1 Proof of Corollary 2.2 In the case of an uncoloured Erdös-Renyi graph, the function C degenerates to a constant c, $L^2 = |E|/n \in [0, \infty)$ and $M = D \in \mathcal{M}(\mathbb{N} \cup \{0\})$. Theorem 2.1 and the contraction principle imply a large deviation principle for D with good rate function

$$\delta(d) = \inf \left\{ J(x, d) \colon x \geq 0 \right\} = \inf \left\{ H(d \,\|\, q_x) + \tfrac{1}{2} x \log x - \tfrac{1}{2} x \log c + \tfrac{1}{2} c - \tfrac{1}{2} x \colon \langle d \rangle \leq x \right\},$$

where q_x is the Poisson distribution with parameter x. We denote by $\delta^x(d)$ the expression inside the infimum and consider the cases (i) $\langle d \rangle \geq c$ and (ii) $\langle d \rangle \leq c$ separately.

Case (i): For any $\varepsilon > 0$, we have

$$\delta^{\langle d \rangle + \varepsilon}(d) - \delta^{\langle d \rangle}(d) = \tfrac{\varepsilon}{2} + \tfrac{\langle d \rangle - \varepsilon}{2} \log \tfrac{\langle d \rangle}{\langle d \rangle + \varepsilon} + \tfrac{\varepsilon}{2} \log \tfrac{\langle d \rangle}{c} \geq \tfrac{\varepsilon}{2} + \tfrac{\langle d \rangle - \varepsilon}{2} \left(\tfrac{-\varepsilon}{\langle d \rangle} \right) + \tfrac{\varepsilon}{2} \log \tfrac{\langle d \rangle}{c} > 0,$$

so that the minimum is attained at $x = \langle d \rangle$.

Case (ii): Under our condition the equation $x = c e^{-2(1 - \langle d \rangle / x)}$ has a unique solution, which satisfies $x \geq \langle d \rangle$. Elementary calculus shows that the global minimum of $y \mapsto \delta^y(d)$ on $(0, \infty)$ is attained at the value $y = x$, where x is the solution of our equation.

6.2 Proof of Corollary 2.4 We begin the proof by defining the continuous linear map $W \colon \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \to [0, \infty)$ by $W(\omega, \varpi) = \tfrac{1}{2} \|\varpi\|$. We infer from Theorem 2.3 and the contraction principle that $W(L^1, L^2) = |E|/n$ satisfies a large deviation principle in $[0, \infty)$ with the good rate function

$$\zeta(x) = \inf \left\{ I(\omega, \varpi) \colon W(\omega, \varpi) = x \right\}.$$

To obtain the form of the rate in the corollary, the infimum is reformulated as unconstrained optimisation problem (by normalising ϖ)

$$\inf_{\substack{\varpi \in \mathcal{M}(\mathcal{X} \times \mathcal{X}) \\ \omega \in \mathcal{M}(\mathcal{X})}} \left\{ H(\omega \,\|\, \mu) + x H(\varpi \,\|\, C\omega \otimes \omega) + x \log 2x + \tfrac{1}{2} \|C\omega \otimes \omega\| - x \right\}. \tag{6.1}$$

By Jensen's inequality $H(\varpi \,\|\, C\omega \otimes \omega) \geq -\log \|C\omega \otimes \omega\|$, with equality if $\varpi = \frac{C\omega \otimes \omega}{\|C\omega \otimes \omega\|}$, and hence, by symmetry of C we have

$$\min_{\varpi \in \mathcal{M}(\mathcal{X} \times \mathcal{X})} \left\{ H(\omega \,\|\, \mu) + x H(\varpi \,\|\, C\omega \otimes \omega) + x \log 2x + \tfrac{1}{2} \|C\omega \otimes \omega\| - x \right\}$$

$$= H(\omega \,\|\, \mu) - x \log \|C\omega \otimes \omega\| + x \log 2x + \tfrac{1}{2} \|C\omega \otimes \omega\| - x.$$

The form given in Corollary 2.4 follows by defining $y = \sum_{a,b \in \mathcal{X}} C(a, b) \omega(a) \omega(b)$.

Bibliography

[Bi04]   J.D. BIGGINS. Large deviations for mixtures. *El. Comm. Probab.* **9** 60–71 (2004).

[BP03]   J.D. BIGGINS and D.B. PENMAN. Large deviations in randomly coloured random graphs. *Preprint 537/03*, Department of Probability and Statistics, University of Sheffield (2003).

[BBL04]   S. BOUCHERON, O. BOUSQUET and G. LUGOSI. Concentration inequalities. In: *Advanced lectures in machine learning*, Eds: O. Bousquet, U. v. Luxburg and G. Rötsch. Springer (2004) 208-240.

[BGL02]   S. BOUCHERON, F. GAMBOA and C. LEONARD. Bins and balls: Large deviations of the empirical occupancy process. *Ann. Appl. Probab.* **12** 607-636 (2002).

[CP03]   C. CANNINGS and D.B. PENMAN. Models of random graphs and their applications. In: *Handbook of Statistics 21. Stochastic Processes: Modelling and Simulation.* Eds: D.N. Shanbhag and C.R. Rao. Elsevier (2003) 51-91.

[DZ98]   A. DEMBO and O. ZEITOUNI. Large deviations techniques and applications. Springer, New York, (1998).

[DA06a]   K. DOKU-AMPONSAH. Large deviations and information theory for hierarchical and networked data structures. PhD Thesis, Bath. (In preparation).

[DA06b]   K. DOKU-AMPONSAH. Asymptotic equipartition principles for simple hierarchical and networked structures. (In preparation).

[Fel67]   W. FELLER. An introduction to probability theory and its applications. Vol. I, Wiley, New York. Third edition, (1967).

[dH00]   F. DEN HOLLANDER. Large deviations. AMS Fields Institute Monographs (2000).

[Pe98]   D.B. PENMAN. Random graphs with correlation structure. PhD Thesis, Sheffield 1998.